Binding of Small Peptide Ligands to Protein Domains Studied by Molecular Dynamics Simulations

Dissertation
zur
Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich

von
Sandra Steiner
aus
Schänis SG

Promotionskomitee

Prof. Dr. Amedeo Caflisch (Vorsitz)
Prof. Dr. Peter Hamm

Zürich 2013

Summary

In order to perform most of their functions in cells, proteins have to assemble to complexes. Specific recognition and binding of partner proteins is a highly complex process and provides thus an interesting research object. In this thesis, we present several approaches of how the knowledge about peptide ligand binding to proteins can be extended by molecular dynamics (MD) simulation studies. Our investigations are focused on two protein families, the PDZ domains and the bromodomains. While the former are involved in signal-transduction complexes in the neuronal tissue, the latter play an important role in epigenetic inheritance, by recognizing acetylated-lysine side chains on histones.

In order to study binding, it is important to study the intrinsic structural and dynamic properties of the protein binding site to which a ligand is bound. Employing MD simulations we can discover different conformations of the protein binding site in the ligand free state, identify key residues and investigate their dynamic and structural characteristics. In particular, we show that a PDZ domain can adopt several conformations in its ligand-free state, while in the ligand-bound state only one of them is accessible. This finding suggests *conformational selection* as the most probable binding mechanism. Simulations of 20 different bromodomains on the other hand show wide variability in biding site accessibility among bromodomains. In particular, for some bromodomains, key residues can interact with each other in such a way that the binding site adopts an obstructed conformation.

Furthermore, we show how MD simulations provide a tool to study unbinding and to detect critical intermolecular interactions and metastable binding poses for a PDZ domain and its peptide ligand. The same PDZ domain is utilized in a study using ultrafast IR spectroscopy, NMR spectroscopy and MD simulations conducted by the group of Prof. Dr. Peter Hamm at the University of Zurich in order to study the conformational changes undergone by the domain upon ligand binding. An azobenzene-photoswitch across the binding groove is employed to mimic this conformational change, and the results propose that the propagation of the perturbation of the water around the binding site is a potential example of dynamic allostery.

A detailed picture of a binding process is obtained by simulating bromodomains together with cosolvent molecules (dimethylsulfoxide, ethanol and methanol) for which the binding rates and the affinities can be extracted. Additionally, theses simulations show that structured water molecules inside the binding cleft can be replaced by cosolvent. Moreover, we present simulations of a bromodomain binding its natural ligand acetyl-lysine which reveal a new, alternative binding pose and enable us to extract the complete kinetics of the binding process.

Summing up, MD simulations are able to unveil atomic-level details of ligand binding and add another (dynamic) dimension to experimental results.

Zusammenfassung

Um ihre Funktion in Zellen zu erfüllen, müssen sich Proteine zu Komplexen zusammenschliessen. Der spezifische Erkennungs- und Bindungsprozess von Proteinpartnern ist ein höchst komplexer Prozess und bietet sich deswegen als interessantes Forschungsobjekt an. In dieser Dissertation präsentieren wir verschiedene Möglichkeiten, um das bekannte Wissen über Bindung von peptidischen Liganden an Proteine mittels Moleküldynamik (MD) Simulationen zu erweitern. Unsere Forschung konzentriert sich auf zwei Protein Familien, die PDZ Domänen, welche in Signalübertragungskomplexen im Nervengewebe involviert sind, und die Bromodomänen, die eine wichtige Rolle spielen in epigenetischer Vererbung, indem sie acetylierte Lysin Seitenketten auf Histon Proteinen erkennen und binden.

Um Bindungprozesse zu erforschen, ist es wichtig die intrinsischen strukturellen und dynamischen Eigenschaften von Proteinbindungsstellen zu untersuchen. Mit Hilfe von MD Simulationen können wir verschiedene Konformationen der Bindungsstelle im Ligand-freien Zustand beobachten, Schlüsselresiduen für den Bindungsprozess identifizieren und deren dynamische und strukturelle Eigenschaften erforschen. Im Speziellen zeigen wir, dass eine PDZ Domäne im Ligand-freien Zustand verschiedene mögliche Konformationen zeigt, während sie im gebundenen Zustand nur eine aufweist. Die Resultate dieser Arbeit weisen auf den Bindungsmechanismus *conformational selection* hin. Weiter zeigen Simulationen von 20 verschiendenen Bromodomänen, dass die Zugänglichkeit der Bindungstelle sehr variabel ist unter den einzelnen Mitgliedern der Familie der Bromodomänen. Einige der Bromodomänen weisen gar eine Konformation auf, in welcher die

Bindungsstelle durch die Interaktion von Schlüsselresiduen komplett blockiert wird.

Ebenfalls zeigen wir hier, dass MD Simulationen angewandt werden können, um das Ablösen von Liganden zu erforschen, und dass damit für eine PDZ Domäne und ihren Liganden kritische intermolekuläre Interaktionen und metastabile Bindungsposen gefunden werden können. Dieselbe PDZ Domäne wurde in einer Studie von der Gruppe von Prof. Dr. Peter Hamm an der Universität Zürich benutzt. In dieser Studie wird mittels ultraschneller IR Spektroskopie, NMR Spektroskopie und MD Simulationen die Konformationsänderung, welche ein Protein bei Ligandbindung durchläuft, untersucht. Um die Konformationsänderung zu imitieren, wird ein photoschaltbarer Cross-Linker über die Bindungsseite gespannt. Die Resultate deuten an, dass die Ausbreitung der Störungen im Wasser rund um die Bindungsstelle ein mögliches Beispiel für dynamische Allosterie ist.

Ein detailliertes Bild von einem Bindungsprozess wird durch Simulationen von Bromodomänen mit Kosolvent-Molekülen (Dimethylsulfoxid, Ethanol und Methanol) gegeben. In diesen Fall können Bindungsraten und -affinitäten anhand der Daten berechnet werden. Weiter zeigen die Simulationen, dass strukturierte Wassermoleküle in der Bindungsstelle durch Kosolvent-Moleküle ersetzt werden können. Darüber hinaus präsentieren wir Simulationen von einer Bromodomäne, welche ihren natürlichen Liganden, acetyliertes Lysin, bindet. Diese Simulationen enthüllen eine alternative Bindungspose und machen es möglich die Kinetik des Bindungsprozesses zu erfassen.

Zusammenfassend können MD Simulationen atomare Details von Bindungsprozessen sichtbar machen und fügen so den experimentellen Resultaten eine weitere (dynamische) Dimension an.

Contents

Chapter 1

Introduction

This work aims to shed light on protein-protein assembly by studying the interactions of proteins and respective mimicking peptides by molecular dynamics (MD) simulations. Thus, the introduction will give an overview over basic concepts of ligand binding, the target protein families used in the studies and the methods employed.

1.1 Ligand-binding of proteins

Proteins are biomolecules fundamental for cell function and thus for life. They constitute not only a major fraction of the mass of all organisms [1] but are responsible for practically all important processes in the cell, e.g., they provide the structural skeleton of the cell, handle ion exchange with the extracellular tissue, play an important role in signal transduction, catalyze chemical reactions, transport metabolites, act as receptors on the cell membrane etc. To perform some of these functions proteins have to assemble to complexes, for others they have to bind metabolites or for example DNA. Binding of specific partners of diverse nature is therefore crucial for protein function. While for biomedical research it is often important to study binding of small organic molecules to proteins, the focus of this thesis is on association of protein domains. However, limited computational power forces us to replace one protein domain by a short peptidic sequence which includes the residues responsible for binding to the other protein domain.

Nevertheless, an understanding of binding of small peptide ligands by protein domains can give partial answers to the question of how proteins assemble.

Binding mechanism. Historically the specificity of molecular recognition was first considered in the end of the 19th century in the context of enzyme reactions by Emil Fischer. He proposed the famous *lock-and-key* model which explained binding specificity with the chemical and structural complementarity of enzyme and substrate, thus the substrate figuratively fits into the enzyme as a key into a lock [2]. This model of rigid structures has several apparent flaws and raises for example the questions why some enzymes are highly selective whereas others can accommodate several structurally different substrates [3]. An attempt towards a more realistic model is the *induced-fit* binding mechanism which has been proposed by Daniel Koshland in 1958 [4]. The induced fit binding mechanism suggests a model where the interaction between ligand and protein binding partner induces a conformational change in the protein. Thus protein and ligand still are "a perfect match" but only after the ligand has induced the appropriate structural changes in the protein. Later on, in the late 1990s, kinetic considerations and the common notion of the protein-folding funnel have led to another model for binding, the *conformational selection* binding mechanism [3, 5]. Proteins are considered to exist in their native state as a dynamical ensemble of inter-converting conformations at the bottom of the folding funnel whose populations follow statistical distributions. In this context the conformational selection model states that a ligand selects and binds to the most suitable conformer of the dynamical ensemble and shifts the equilibrium in favor of the bound conformation. The thermodynamic cycle shown in Figure 1.1 depicts the differences between an induced-fit and a conformational selection binding mechanism. The former corresponds to conduction of reaction 2 after reaction 1, while the latter corresponds to reaction 3 followed by reaction 4 [3].

For all the three binding mechanisms experimental evidence has been found, for example [6, 7, 8, 9, 10]. In fact, the relatively new view of proteins having dynamical structures makes the conformational selection model very appealing, although it doesn't

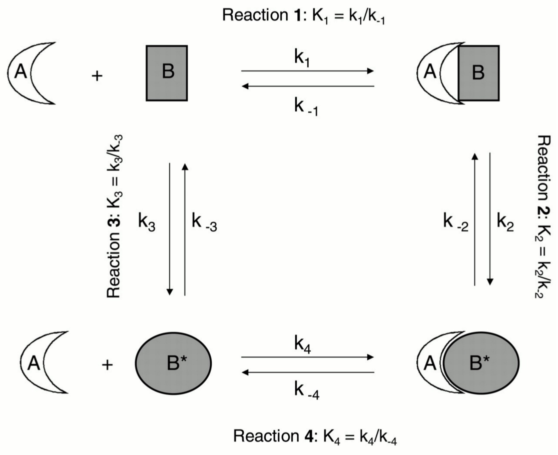


Figure 1.1: **Schematic illustration of the reaction of molecules A and B to the complex AB*.** Binding by induced-fit corresponds to forming the unstable complex AB followed by a conformational change from B to B*. On the other hand if A selects directly the conformer B* and forms AB*, binding proceeds by conformational selection. The figure was taken from [3].

necessarily contradict the lock-and-key and induced-fit models. If for example a molecule has a very stable binding site with very little flexibility, the free energy landscape of this binding site is smooth with one clear minimum. A potential ligand can either perfectly match the binding site or induce small structural changes in it, which although unfavorable for the protein itself, may be compensated by protein-ligand enthalpic contributions. In order to reconcile the three models it has been suggested that the most appropriate way to describe the binding mechanism between proteins and their respective ligands is a combination of conformational selection and induced-fit. Thus, a ligand first selects the best matching conformation of the protein and upon association minor structural readjustments are conducted [3, 11]. Computational evidence for such a combination has been found [12]. Furthermore, if it is left open to which extent conformational preselection of the conformer and following structural readjustment take place, the three originally proposed mechanisms can naturally be seen as the extreme cases of such a combination.

1.1.1 PDZ domains

PDZ domains are so-called protein-protein interaction domains and appear often in arrays and together with other interaction domains in multidomain scaffold proteins. By their specific interaction with (mostly) the C-terminal segments of target proteins they can assemble large molecular complexes [13, 14]. These complexes regulate diverse biological activities, for example in the neuronal tissue, particularly in the excitatory synapses where post-synaptic density proteins contain many PDZ domains, they regulate signal transduction.

PDZ domains are small, consist of ≈ 90 residues and have a highly conserved fold. They usually consist of five to six β-strands ($\beta 1 - \beta 6$) and two α-helices ($\alpha 1, \alpha 2$), but they may differ among each other by additional structural elements, lengths of loops, β-strands and helices. Such an additional structural element is for example the third α-helix at the C-terminus of PDZ3 of postsynaptic density protein PSD-95.

The C-terminus of the ligand protein binds in an extended conformation to an elongated pocket between the $\beta 2$ strand and the $\alpha 2$ helix thereby augmenting the antiparallel $\beta 2$-$\beta 3$ hairpin into a three-stranded β-sheet (see Figure 1.2).

The diverse biofunctions of PDZ domains but also their conserved fold and specificity for peptide-motifs make them a very interesting research target. For several PDZ domains the binding mechanism was experimentally investigated, yielding evidence for induced-fit and lock-and-key binding mechanisms for different PDZ domains. However, our MD simulations suggest that conformational selection is also a possible mechanism for PDZ domains, as it has been observed for the third PDZ domain of postsynaptic density protein PSD-95 (see Chapter 2).

Furthermore, many PDZ domains have been shown to possess allosteric properties, see for example [15, 16, 17, 18, 19]. Since large conformational changes and rearrangements are rare among the resolved structures, the research on allostery in PDZ domains focuses on dynamic allostery, e.g., the question how the signal of ligand binding is propagated dynamically through the domain, e.g., to the C-terminal end of the domain and thus to other domains in the protein. Therefore, the intra-molecular signaling properties have

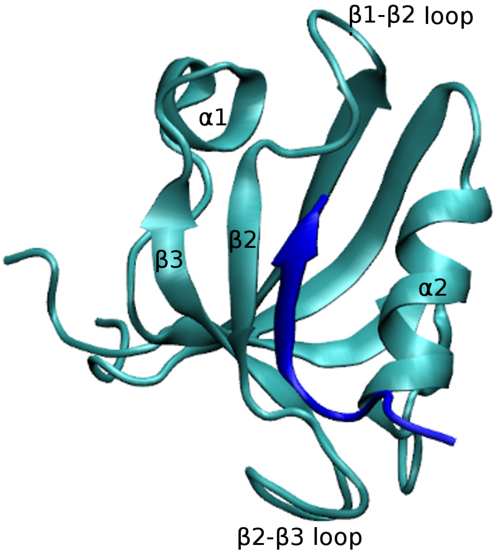


Figure 1.2: **Ribbon illustration of the NMR structure of the complex between the second PDZ domain of human phosphatase hPTP1E (PDZ2) and a C-terminal peptide from the guanine nucleotide exchange factor RA-GEF-2 (PDB code 1D5G).** The ligand-peptide (blue) extends the β2-β3 hairpin of the domain (cyan) into a three-stranded β-sheet.

been of particular interest. It has for example been shown that evolutionary conserved signaling pathways seem to connect the binding site residues to distal parts of the PDZ fold [20]. Furthermore, computational studies yielded similar pathways by investigating residue-residue interaction energy correlations [21]. An NMR spectroscopy study has shown that the additional structural element of PDZ3, a C-terminal α-helix (α3) distal to the binding site, influences ligand affinity by modifying the side chain dynamics of the whole domain [18]. See Chapter 4 for how a signal can be dynamically propagated in the case of the second PDZ domain of human phosphatase hPTP1E.

1.1.2 Bromodomains

Bromodomains are so-called reader domains involved in epigenetic processes. The concept of epigenetics were first introduced by Waddington in 1939 [22] and deals with the study of heritable changes of the phenotype without changing the genotype, i.e., changes in gene expression without changing the underlying DNA-sequence. Posttranslational modifications (PTMs), for example acetylation or methylation of residues on histones, potentially represent a method for epigenetic inheritance. Investigation of such histone modifications led to the notion of “writer”-, “reader”- and “eraser”-domains, which mean protein families that are able to add (write), recognize (read) and remove (erase) PTMs. Bromodomains belong to the “reader”-domains and recognize acetylated lysine on histones by binding to it.

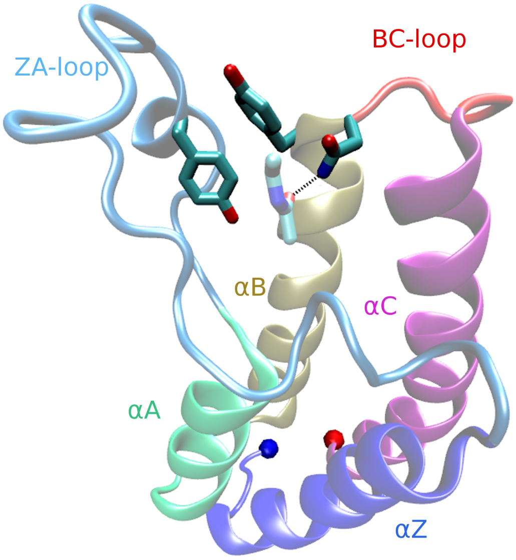


Figure 1.3: **Ribbon illustration of the crystal structure of the complex between CREBBP and acetyl-lysine (PDB code 3P1C).** Each of the four α-helices and two binding site loops is displayed with a different color. The side chains of the conserved Tyr1125 of the ZA-loop, and Tyr1167-Asn1168 of the BC-loop are emphasized (sticks) together with the side chain of acetyl-lysine (sticks, light colors). The characteristic hydrogen bond between the conserved Asn of CREBBP and acetyl-lysine is shown as dashed line. The N-terminus and C-terminus are shown with a blue and red sphere, respectively.

They are modules of about 110 residues which show a conserved four-helix bundle topology (Figure 1.3) in which the ZA-loop and BC-loop connect the first two α helices (called Z and A) and last two α helices (called B and C), respectively [23, 24]. The acetyl-lysine binding site is very similar in all structures of bromodomains [23]. The BC-loop contains the evolutionary conserved Asn side chain [25] which acts as hydrogen bond donor to the acetylated lysine side chain.

Bromodomains are of particular biomedical interest due to their potential role in tumors and inflammation [26, 27, 28]. The development of small molecules which disrupt the interaction of bromodomains with actyl-lysine on histones is thus of great interest. Remarkably, an inhibitor for BET bromodomains to treat NUT midline carcinoma is already in phase I/II clinical trial [29].

1.2 Computational methods and analysis tools

In order to study problems like ligand-binding to proteins or protein folding we can not only apply experimental but also computational tools. In this thesis we have investigated binding/unbinding of peptide ligands to protein domains and structural as well as dynamical properties of the protein binding sites with MD simulations. This method simulates the physical evolution in time of systems by allowing their atoms to interact over a fixed time period and employing the classical Newtonian equation of motion to determine the position of each atom at every time step. Thus, in this way we can simulate the evolution in time of protein domains, protein/ligand complexes, or even of the process of ligand binding. The result is an atomic description of processes in time. The first biological problem to which MD simulations were applied was the study of the dynamics of a folded globular protein, i.e., the bovine pancreatic trypsin inhibitor (BPTI) in 1977 [30]; that trajectory was only 9 ns long. Since then the computational power has grown much, thus today processes on the 1-100 μs timescale can be studied by all-atom simulations.

Naturally it is not enough just to simulate such a process. In fact, the quantitative and qualitative analysis of the simulations, that is of this atomic descriptions, is the crucial and difficult part of our work as computational scientists. From simulation trajectories time evolution or averages over time of quantities like radius of gyration, internal energy, secondary structure propensity, interaction energy and root mean square deviation (RMSD) to a reference structure can be readily computed. A more sophisticated approach is to calculate the projection of the free energy surface onto useful coordinate(s). How this can be employed to investigate the plasticity of protein binding sites or for studying the ligand binding process can be seen in Chapters 2 and 7. The theoretical background of MD simulations is given in the next subsection, followed by a description of how to project the free energy surface in a barrier-preserving way.

1.2.1 MD simulations

In MD simulations based on classical mechanics, atoms of a system interact (virtually) over a specified amount of time and the evolution of the system in time is obtained by numerically solving Newton's equation of motion (1.1) for each atom i in an iterative way:

$$m_i \frac{d^2}{dt^2}\mathbf{r_i} = m_i \mathbf{a_i} = \mathbf{F_i} = -\frac{\partial}{\partial \mathbf{r_i}} E(\mathbf{r}) \tag{1.1}$$

where $\mathbf{r}$ are the coordinates of all atoms in the system, $\mathbf{r_i}$ are the coordinates of atom i, m_i is the mass of atom i, $\mathbf{F_i}$ the force acting on it and E is the potential energy of the system.

The energy of the system is calculated using an empirical energy potential which is usually calculated from two contributions, i.e., the covalent and the non-covalent contributions: $E(\mathbf{r}) = E_{cov}(\mathbf{r}) + E_{noncov}(\mathbf{r})$. The covalent term is the sum over the energy of bonds b, angles θ, dihedral ϕ and impropers ω:

$$\begin{aligned} E_{cov}(\mathbf{r}) &= E_b + E_\theta + E_\phi + E_\omega \\ &= \sum_b \frac{1}{2}k_b(b-b_0)^2 + \sum_\theta \frac{1}{2}k_\theta(\theta-\theta_0)^2 + \\ &\quad \sum_\phi k_\phi[1+cos(n\phi-\delta)] + \sum_\omega \frac{1}{2}k_\omega(\omega-\omega_0)^2 \end{aligned} \tag{1.2}$$

where k_b, k_θ, k_ϕ and k_ω are the empirically determined force constants modeling the rigidity of bonds, angles, dihedrals and impropers and b_0, θ_0, ω_0 are the respective equilibrium values for bond length, angle width and improper. Furthermore for E_ϕ, the parameters n and δ determine the number and location of energy minima.

The non-covalent term is computed as the sum of van der Waals and electrostatic energy, i.e., of the Lennard-Jones and the Coulombic potential.

$$\begin{aligned} E_{noncov}(\mathbf{r}) &= E_{LJ} + E_C \\ &= \sum_{i\neq j} \epsilon_{ij} \left[\left(\frac{r_{ij}^{min}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}^{min}}{r_{ij}} \right)^{6} \right] + \sum_{i\neq j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \end{aligned} \tag{1.3}$$

where r_{ij}^{min} is the equilibrium Lennard-Jones separation of atoms i and j and ϵ_{ij} is such that $E_{LJ}(r_{ij}^{min}) = -\epsilon_{ij}$. For E_C, we need the partial charges q_i, q_j of atoms i, j and the vacuum permittivity ϵ_0.

1.2.2 Cut-based one-dimensional free energy profiles and optimization of the reaction coordinate

In order to construct a one-dimensional free energy profile the MD trajectory has to be further processed. There are essentially two ways to do so, by grouping the snapshots with a clustering algorithm or by binning along a reaction coordinate (RC). The binning suggests the conventional, thermodynamically inspired way of projecting the free energy as $F_H(i) = -kT \ln Z_H(i)$ where k is the Boltzmann constant, T is the temperature, $Z_H(i)$ denotes the partition function which is equal to the density of bin i, i.e., $Z_H(i) = \frac{\#\text{frames in bin } i}{\text{bin size}}$ and F_H is short for histogram-based free energy. A more kinetically motivated approach that is able to preserve the free energy barriers is to compute the cut-based free energy profile (cFEP) [31, 32, 33].

Cut-based free energy profile. This approach emulates the cuts in flow-networks. Based on the simulation data a network of system configurations is generated. The nodes of this network are the clusters of snapshots obtained either by applying a clustering algorithm or by binning along an RC. The links between them correspond to transitions along the trajectory between them and are weighted according to their frequency in the trajectory [34]. The partition function of the free energy barrier separating two nodes i and j is equal to the value of minimum cut between the nodes in the network [31]. The minimum cut of a network with respect to nodes i and j is the partition into two sets A

and B, with $i \in A$ and $j \in B$, such that the sum of weights of links between A and B is minimal. The value of minimum cut is then the sum of weights of links and corresponds to the maximal flow from i to j and thus to the free energy barrier separating two nodes i and j.

Since computing $2^{\#nodes}$ minimum cuts is computationally expensive and after all yields an information which is not projectable onto one coordinate, in practice an approximation for the value of minimum cut is used such that the free energy can be projected along one coordinate. This is done in the following way: For each node, the set of nodes is partitioned into two groups A and B using the mean first passage time (MFPT) to a reference node as an order parameter. The free energy barrier between sets A and B is calculated as $G = -kT \ln(Z_{AB}/Z)$, where Z_{AB}/Z is the relative partition function which represents the statistical weight of the transitions between sets A and B (for further details see [31, 33]). The result is a one-dimensional profile along the reaction coordinate Z_A/Z (i.e., the relative partition function representing the statistical weight of set A) which preserves the barrier height between the free energy basins. It has been shown that the barrier height between two basins approximates the minimum cut between two reference nodes i and j of the individual basins with high accuracy [33]. Figure 1.4 illustrates the network representation of a clusterized MD trajectory, the free energy basins as parts of the network with high connectivity and the resulting one-dimensional cFEP.

Similarly, using a classification of each snapshot according to an RC instead of applying a clustering algorithm, one can estimate in continuous terms the statistical weight of the transitions through each value x of the RC and thereby the cut-based partition function $Z_C(x)$ at a point x [32] (C stands for cut-based as opposed to H for histogram-based). Accordingly, the computation of the cut-based free energy is performed as $F_C(x) = -kT \ln(Z_C(x))$. Naturally, one could also use the relative partition function $Z_C(x)/Z$ as proposed above to calculate the free energy. Since this induces just a shift of the profile, the computed barriers from the $F_C(x) = -kT \ln(Z_C(x))$ and $F_C^{'}(x) = -kT \ln(Z_C(x)/Z)$ are the same. In terms of bins the following formula

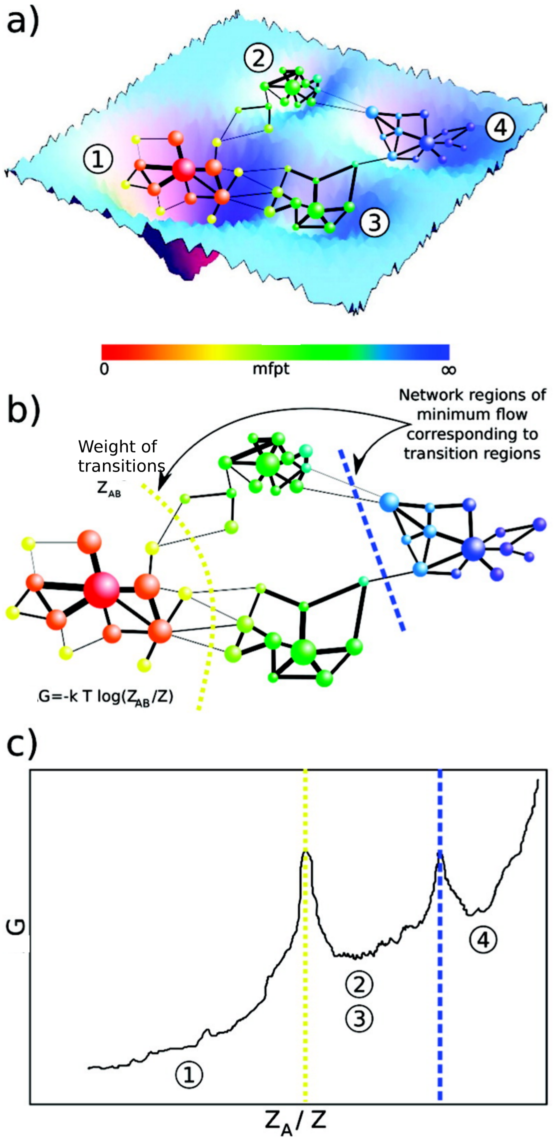


Figure 1.4: **Schematic illustration of cFEP upon clustering.** (a) The MD trajectory is coarse-grained into nodes of the network by clustering and two nodes are linked if a direct transition between them is observed in the trajectory. The MFPT to the reference node is calculated for each node analytically. (b) For each value of MFPT, the network is partitioned into sets A and B, where A consists of all nodes with lower MFPT and B are the remaining nodes. The free energy G of the barrier between the two sets A and B is equal to the statistical weight of the transitions between them. (c) The cFEP is the projection of the free energy onto the relative partition function Z_A/Z. For each value of MFTP, the point $(Z_A/Z, -kT\ln(Z_{AB}/Z))$ is added to the cFEP. Up to the first barrier the free energy is projected without overlap (basin 1), but after the first barrier, two or more basins may overlap (e.g. here, basins 2 and 3) if they have the same kinetic distance from the reference node. Figure and legend taken from [35] and slightly adapted.

applies $Z_C(i, i+1) = \frac{1}{2}(\sum_{j \le i < k} n_{jk} + \sum_{j > i \ge k} n_{jk})$ where n_{ij} is the number of transitions from bin j to bin i. In terms of continuous RC the formula becomes $Z_C(x) = \frac{1}{2}\sum_t \theta((x(t) - x)(x - x(t + \Delta t)))$ where θ is the Heaviside step function and Δt is the sampling interval. A schematic illustration of this procedure is shown in Figure 1.5.

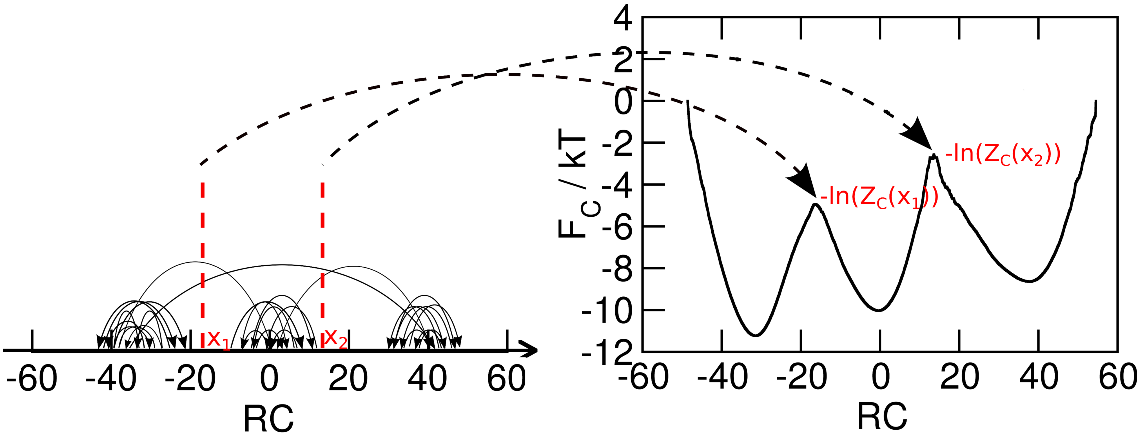


Figure 1.5: **Schematic illustration of cFEP upon binning.** By computing its RC value (real number), every MD snapshot can be uniquely placed on the line representing the real numbers. The transitions are given naturally from snapshot i to snapshot $i + 1$ and are here illustrated by curved arrows (left). Then for a fixed RC value, e.g. x_1, the cut-based partition function $Z_C(x_1)$ is the number of transitions through this point, i.e., the number of curved arrows crossing the dashed line at x_1. Computing this values for every x gives the cFEP as $F_C(x) = -kT \ln(Z_C(x))$ (right).

Whether an RC is adequate or not depends usually on the system one is investigating. Geometrically inspired RCs such as the number of native contacts, the radius of gyration or the root mean square deviation from a starting structure are often employed. However, they have often been shown to suffer from overlap and lead thus to underestimation of barrier height and an inaccurate description of the underlying dynamics [31]. A way to find a "good" reaction coordinate, was developed by Krivov [36]. The aim of this optimization procedure is to find an appropriate coordinate on which the dynamics are diffusive at the barrier regions. This is achieved by optimizing a linear functional which is based on the idea of "native contacts". We used this approach to investigate the free energy surface of ligand-bound and ligand-free PDZ3 domain in Chapter 2.

1.3 Structure of the thesis

From a computational point of view different aspects of ligand binding are interesting and can be studied. Most naturally one of them is the binding mechanism. Here, this question has been addressed for the third PDZ domain of the synaptic protein PSD-95 (PDZ3). Chapter 2 explains how the binding mechanism can be retrospectively determined by simulations of the ligand-free and ligand-bound forms of PDZ3. Furthermore, one can investigate unbinding of peptide ligands and search for the most stable intermolecular interactions. This has been done for PDZ2 of human tyrosine phosphatase PTPL1 in Chapter 3. For the same domain also a combined experimental and computational study has been conducted making use of a photoswitchable linker in order to emulate conformational change upon peptide binding (Chapter 4). Moreover, since MD simulations render information at atomic resolution, one can study the properties of the binding site itself in atomic detail, i.e., its structural plasticity, (dynamical) properties of its key residues or the role of structured water molecules in the binding site, which are able to coordinate interactions between protein and ligand. Studying such properties is especially important for protein domains involved in development of diseases that can be targeted by inhibitor-molecules. For the family of bromodomains such research has been conducted and is presented in Chapters 5 and 6. Finally, currently available computational power allows tracking of the complete binding process of small peptides. In Chapter 7 the binding kinetics and binding poses of an acetyl-lysine mimicking monopeptide to bromodomain TAF1(2) are extracted from MD simulations.

Chapters 2, 4, 5, 6 and 7 of the thesis correspond to different publications or submitted manuscripts. Chapter 3 is an excerpt of ongoing research.

- **Chapter 2: Peptide binding to the PDZ3 domain by conformational selection.**
- **Chapter 3: Peptide unbinding from PDZ2.**

- **Chapter 4: Kinetic response of a photo-perturbed allosteric protein.**
- **Chapter 5: Does bromodomain flexibility influence histone recognition?**
- **Chapter 6: Structured water molecules in bromodomains binding site can be displaced by cosolvent.**
- **Chapter 7: Mechanism and kinetics of acetyl-lysine binding to bromodomains.**

Bibliography

[1] Donald Voet and Judith G Voet. *Biochemistry.* John Wiley and Sons, Inc., 1995.

[2] E Fischer. Einfluss der Configuration auf die Wirkung der Enzyme. *Ber Dtsch Chem Ges*, 27:2984–2993, 1894.

[3] Hans Rudolf Bosshard. Molecular recognition by induced fit: How fit is the concept? *Physiology*, 16(4):171–173, 2001.

[4] D. E. Jr. Koshland. Application of a theory of enzyme specificity to protein synthesis. *Proc Natl Acad Sci U S A*, 44:98–104, 1958.

[5] Buyong Ma, Sandeep Kumar, Chung-Jung Tsai, and Ruth Nussinov. Folding funnels and binding mechanisms. *Protein Eng.*, 12(9):713–720, 1999.

[6] B.C. Braden, B.A. Fields, X. Ysern, F.A. Goldbaum, W. Dall'Acqua, F.P. Schwarz, R.J. Poljak, and R.A. Mariuzza. Crystal structure of the complex of the variable domain of antibody D1.3 and turkey egg white lysozyme: A novel conformational change in antibody CDR-L3 selects for antigen. *J Mol Biol*, 257(5):889–894, 1996.

[7] Jun Zhang, Paul J Sapienza, Hengming Ke, Aram Chang, Sarah R Hengel, Huanchen Wang, George N Phillips, and Andrew L Lee. Crystallographic and nuclear magnetic resonance evaluation of the impact of peptide binding to the second PDZ domain of protein tyrosine phosphatase 1e. *Biochemistry*, 49(43):9280–9291, 2010.

[8] Stefano Gianni, Tine Walma, Alessandro Arcovito, Nicoletta Calosci, Andrea Bellelli, Ake Engström, Carlo Travaglini-Allocatelli, Maurizio Brunori, Per Jemth, and

Geerten W Vuister. Demonstration of long-range interactions in a PDZ domain by NMR, kinetics, and protein engineering. *Structure*, 14(12):1801–1809, 2006.

[9] Lukas Leder, Christine Berger, Susanne Bornhauser, Hans Wendt, Friederike Ackermann, Ilian Jelesarov, and Hans Rudolf Bosshard. Spectroscopic, calorimetric, and kinetic demonstration of conformational adaptation in peptide-antibody recognition. *Biochemistry*, 34(50):16509–16518, 1995.

[10] Dariusz Ekonomiuk and Amedeo Caflisch. Activation of the west nile virus NS3 protease: molecular dynamics evidence for a conformational selection mechanism. *Protein Sci*, 18(5):1003–1011, 2009.

[11] Peter Csermely, Robin Palotai, and Ruth Nussinov. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci*, 35(10):539–546, 2010.

[12] Daniel-Adriano Silva, Gregory R Bowman, Alejandro Sosa-Peinado, and Xuhui Huang. A role for both conformational selection and induced fit in ligand binding by the LAO protein. *PLoS Comput Biol*, 7(5):e1002054, 2011.

[13] Eunjoon Kim and Morgan Sheng. PDZ domain proteins of synapses. *Nat Rev Neurosci*, 5(10):771–781, 2004.

[14] Jing-Song Fan and Mingjie Zhang. Signaling complex organization by PDZ domain proteins. *Neurosignals*, 11(6):315–321, 2002.

[15] Lieke C J van den Berk, Elena Landi, Tine Walma, Geerten W Vuister, Luciana Dente, and Wiljan J A J Hendriks. An allosteric intramolecular PDZ-PDZ interaction modulates PTP-BL PDZ2 binding specificity. *Biochemistry*, 46(47):13629–13637, 2007.

[16] Mingjie Zhang. Scaffold proteins as dynamic switches. *Nat Chem Biol*, 3(12):756–757, 2007.

[17] Jungsan Sohn, Robert A Grant, and Robert T Sauer. Allostery is an intrinsic property of the protease domain of DegS: implications for enzyme function and evolution. *J Biol Chem*, 285(44):34039–34047, 2010.

[18] Chad M Petit, Jun Zhang, Paul J Sapienza, Ernesto J Fuentes, and Andrew L Lee. Hidden dynamic allostery in a PDZ domain. *Proc Natl Acad Sci U S A*, 106(43):18249–18254, 2009.

[19] Z. Nevin Gerek and S. Banu Ozkan. Change in allosteric network affects binding affinities of PDZ domains: Analysis through perturbation response scanning. *PLoS Comput Biol*, 7(10):e1002154, 2011.

[20] S. W. Lockless and R. Ranganathan. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science*, 286(5438):295–299, 1999.

[21] Yifei Kong and Martin Karplus. Signaling pathways of PDZ2 domain: a molecular dynamics interaction correlation analysis. *Proteins*, 74(1):145–154, 2009.

[22] C H Waddington. *An introduction to modern genetics.* George Allen Unwin Ltd., 1956 (1939).

[23] Panagis Filippakopoulos, Sarah Picaud, Maria Mangos, Tracy Keates, Jean-Philippe Lambert, Dalia Barsyte-Lovejoy, Ildiko Felletar, Rudolf Volkmer, Susanne Müller, Tony Pawson, Anne-Claude Gingras, Cheryl H. Arrowsmith, and Stefan Knapp. Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell*, 149(1):214–231, 2012.

[24] Lewis R. Vidler, Nathan Brown, Stefan Knapp, and Swen Hoelder. Druggability analysis and structural classification of bromodomain acetyl-lysine binding sites. *J. Med. Chem.*, 55(17):7346–7359, 2012.

[25] David J. Owen, Prisca Ornaghi, Ji-Chun Yang, Nicholas Lowe, Philip R. Evans, Paola Ballario, David Neuhaus, Patrizia Filetici, and Andrew A. Travers. The struc-

tural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase Gcn5p. *EMBO J*, 19(22):6141–6149, 2000.

[26] Panagis Filippakopoulos, Jun Qi, Sarah Picaud, Yao Shen, William B Smith, Oleg Fedorov, Elizabeth M Morse, Tracey Keates, Tyler T Hickman, Ildiko Felletar, Martin Philpott, Shonagh Munro, Michael R McKeown, Yuchuan Wang, Amanda L Christie, Nathan West, Michael J Cameron, Brian Schwartz, Tom D Heightman, Nicholas La Thangue, Christopher A French, Olaf Wiest, Andrew L Kung, Stefan Knapp, and James E Bradner. Selective inhibition of BET bromodomains. *Nature*, 468(7327):1067–1073, 2010.

[27] Johannes Zuber, Junwei Shi, Eric Wang, Amy R. Rappaport, Harald Herrmann, Edward A. Sison, Daniel Magoon, Jun Qi, Katharina Blatt, Mark Wunderlich, Meredith J. Taylor, Christopher Johns, Agustin Chicas, James C. Mulloy, Scott C. Kogan, Patrick Brown, Peter Valent, James E. Bradner, Scott W. Lowe, and Christopher R. Vakoc. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature*, 478(7370):524–528, 2011.

[28] Mark A. Dawson, Tony Kouzarides, and Brian J.P. Huntly. Targeting epigenetic readers in cancer. *N. Engl. J. Med.*, 367(7):647–657, 2012.

[29] David S. Hewings, Timothy P. C. Rooney, Laura E. Jennings, Duncan A. Hay, Christopher J. Schofield, Paul E. Brennan, Stefan Knapp, and Stuart J. Conway. Progress in the development and application of small molecule inhibitors of bromodomain–acetyl-lysine interactions. *J. Med. Chem.*, 55(22):9393–9413, 2012.

[30] J. A. McCammon, B. R. Gelin, and M. Karplus. Dynamics of folded proteins. *Nature*, 267(5612):585–590, 1977.

[31] Sergei V Krivov and Martin Karplus. One-dimensional free-energy profiles of complex systems: progress variables that preserve the barriers. *J Phys Chem B*, 110(25):12689–12698, 2006.

[32] Sergei V Krivov and Martin Karplus. Diffusive reaction dynamics on invariant free energy profiles. *Proc Natl Acad Sci U S A*, 105(37):13841–13846, 2008.

[33] Sergei V. Krivov, Stefanie Muff, Amedeo Caflisch, and Martin Karplus. One-dimensional barrier-preserving free-energy projections of a β-sheet miniprotein: New insights into the folding process. *J. Phys. Chem. B*, 112(29):8701–8714, 2008.

[34] Francesco Rao and Amedeo Caflisch. The protein folding network. *J Mol Biol*, 342(1):299–306, 2004.

[35] Philipp Schuetz, René Wuttke, Benjamin Schuler, and Amedeo Caflisch. Free energy surfaces from single-distance information. *J. Phys. Chem. B*, 114(46):15227–15235, 2010.

[36] Sergei V Krivov. Is protein folding sub-diffusive? *PLoS Comput Biol*, 6(9):e1000921, 2010.

Chapter 2

Peptide binding to the PDZ3 domain by conformational selection

Steiner, S. and Caflisch A. *Proteins: Structure, Function and Bioinformatics*, 80(11): 2562-2572, 2012

PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Peptide binding to the PDZ3 domain by conformational selection

Sandra Steiner and Amedeo Caflisch*

Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

ABSTRACT

The PDZ domains, a large family of peptide recognition proteins, bind to the C-terminal segment of membrane ion channels and receptors thereby mediating their localization. The peptide binding process is not known in detail and seems to differ among different PDZ domains. For the third PDZ domain of the synaptic protein PSD-95 (PDZ3), a lock-and-key mechanism was postulated on the basis of the almost perfect overlap of the crystal structures in the presence and absence of its peptide ligand. Here, peptide binding to PDZ3 is investigated by explicit solvent molecular dynamics (MD) simulations (for a total of 1.3 μs) and the cut-based free energy profile method for determining free energy barriers and basins. The free energy landscape of apo PDZ3 indicates that there are multiple basins within the native state. These basins differ by the relative orientation of the α2 helix and β2 strand, the two secondary structure elements that make up the peptide binding site. Only the structure with the smallest aperture of the binding site is populated in the MD simulations of the complex whose analysis reveals that the peptide ligand binds to PDZ3 by selecting one of three conformations. Thus, the dynamical information obtained by the atomistic simulations increment the static, that is, partial, picture of the PDZ3 binding mechanism based on the X-ray crystallography data. Importantly, the simulation results show for the first time that conformational selection is a possible mechanism of peptide binding by PDZ domains in general.

Proteins 2012; 80:2562–2572.
© 2012 Wiley Periodicals, Inc.

Key words: free energy surface; PDZ domain; molecular dynamics; reaction coordinate optimization.

INTRODUCTION

PDZ domains are found in various proteins with diverse functions either alone or as arrays. They mediate interactions by binding to the C-terminal segment of target proteins and play an important role in signal transduction and scaffolding.[1,2] PDZ domains are small (≈90 residues) and have a highly conserved fold consisting of five to six β-strands (β1–β6) and two α-helices (α1, α2). The C-terminus of the ligand protein binds in an extended conformation to an elongated pocket (termed hereafter binding site groove) between the β2 strand and the α2 helix thereby augmenting the antiparallel β2–β3 hairpin into a three-stranded β-sheet.

A large number of crystal and nuclear magnetic resonance (NMR) structures of PDZ domains in ligand-free (apo) and complexed (holo) state have been solved,[3–6] and several studies have investigated the binding mechanisms of different PDZ domains and their peptide binding partners using a range of experimental techniques.[5–10] Most of them have concerned three of the most thoroughly studied PDZ domains, the second PDZ domain of human tyrosine phosphatase PTPL1 or murine tyrosine phospahtase PTP-BL (human PDZ2, murine PDZ2) and the third PDZ domain of human synaptic protein PSD-95 (PDZ3). In particular, a lock-and-key binding mechanism has been suggested for PDZ3 due to a lack of visible conformational changes when overlaying crystal structures of apo and holo PDZ3.[3] This finding has been corroborated by kinetic experiments suggesting a one-step binding mechanism without rate-limiting conformational changes.[7,8] For murine PDZ2, on the other hand, NMR structures together with ultrafast kinetic experiments provide evidence for an induced-fit binding mechanism.[8] Additionally, also for the human form of PDZ2 inspection of NMR structures suggests an induced-fit binding mechanism.[5] Yet, more recent crystallographic experiments corroborated by residual dipolar couplings

---

Additional Supporting Information may be found in the online version of this article.
Grant sponsor: Swiss National Science Foundation (to A.C.).
*Correspondence to: A. Caflisch Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland. E-mail: caflisch@bioc.uzh.ch
Received 13 April 2012; Revised 4 June 2012; Accepted 14 June 2012
Published online 2 July 2012 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.24137

© 2012 WILEY PERIODICALS, INC.

were able to detect only slight conformational changes in the crystal structures.[6] Complementing the crystal structures with Carr-Purcell-Meiboom-Gill relaxation dispersion data indicated a lock-and-key mechanism to be the most suitable binding mechanism for human PDZ2.[6] An interesting structural feature regarding binding is the β2–β3 loop. It has been reported that for some PDZ domains this loop contributes to protein–ligand interactions[5] and that it has often enhanced conformational flexibility.[4,10] Therefore, it is speculated to assist in ligand-selectivity of PDZ domains.

Several experimental and computational studies have focused on the allosteric behavior of PDZ domains.[11–15] Large conformational changes and rearrangements are rare among the resolved structures, thus particularly the intramolecular signaling properties have been of interest. Evolutionary conserved signaling pathways seem to connect the binding site residues to distal parts of the PDZ fold.[16] Computational studies yielded similar pathways by investigating residue–residue interaction energy correlations.[17] An NMR spectroscopy study has shown that the additional structural element of PDZ3, a C-terminal α-helix (α3), influences ligand affinity by modifying the side chain dynamics of the whole domain.[14] This study has provided evidence of dynamic allostery in the PDZ3 domain, as its C-terminal α-helix is distal to the ligand and dynamical fluctuations throughout the protein cause the changes in binding affinity.[14]

Binding mechanisms of several proteins have been investigated by atomistic simulations. It has been suggested that the distinction between the induced-fit and the conformational selection model is not absolute.[18] A computational study regarding the binding mechanism of Lysine-, Arginine-, Ornithine-binding (LAO) protein corroborates this suggestion as it appears from the simulations that binding by conformational selection is followed by conformational readjustment,[19] thus a combination of the two mechanisms "conformational selection" and "induced-fit" is at work. On the other hand, molecular dynamics (MD) simulations suggest that the flaviviral NS2B-NS3pro protease actually binds its peptide substrate by conformational selection,[20] although crystal structures displaying conformational differences between apo and holo conformations of NS2B-NS3pro had previously led to the speculation of an induced-fit binding mechanism. The MD simulations started from the apo structure showed that even without the substrate the catalytic competent conformation formed spontaneously and was kinetically stable on a 50-ns time scale.[20] Furthermore, a two-state model of the binding free energy surface of PDZ3 has recently emerged from atomistic Monte Carlo simulations with restraints on the protein.[21]

Here, the mechanism of peptide binding by the third PDZ domain (PDZ3) from the human synaptic protein PSD-95 is investigated by explicit solvent MD simulations of its complexed (holo) and peptide-free (apo) state. This simulation study was motivated by the heterogeneity of binding mechanisms suggested by using a variety of experimental methods. In this context, there are several outstanding questions: How to reconcile the lock-and-key mechanism postulated for peptide binding to PDZ3[3,7] with the evidence for a "delocalized conformational entropy mechanism" mediated by the α3 distal helix[14]? Is it possible to shed light on the plasticity of the β2–β3 loop and helix α2 and their influence on the binding mechanism? Is the functional role of α3 only of dynamical nature or has α3 also a structural impact? How can the equilibrium native state ensembles for apo and holo PDZ3 be characterized structurally? The simulation results indicate that PDZ3/peptide association follows most likely a conformational selection mechanism. While the complexed state is rather rigid the peptide-free state displays conformational flexibility. In particular, the β2–β3 loop and the tilting of the helix α2 characterize structurally the different free energy basins of the apo state.

MATERIALS AND METHODS

MD simulations

The explicit solvent MD simulations were carried out with CHARMM[22] and GROMACS[23] using the CHARMM22 all-hydrogen force field[24] and the TIP3P water model.[25] The coordinates of the complex between PDZ3 and the 5-residue C-terminal motif Lys-Gln-Thr-Ser-Val were downloaded from the protein data bank (PDB code 1BE9) and used as starting structures. Note that residues 303–402 of the PDZ3 domain are termed here residues 1–100, in accordance with the work of Petit et al.[14] To reproduce neutral pH conditions, the side chains of aspartates and glutamates as well as the C-terminus of the peptide were negatively charged, those of lysines and arginines were positively charged, and histidines were considered neutral. The protein was immersed in a truncated octahedral box of preequilibrated water molecules. The size of the box was chosen to have a minimal distance of 12 Å between the boundary and any atom of the protein. The simulation box contained five and four potassium ions to compensate for the net negative charge of apo PDZ3 and PDZ3/peptide complex, respectively.

Periodic boundary conditions were applied, long-range electrostatic interactions were treated with the Particle Mesh Ewald method[26] and the van der Waals interactions were truncated at a cutoff of 10 Å using a switch function starting from 8 Å. The MD simulations were performed at constant temperature (298 K) using the Berendsen thermostat[27] and constant pressure (1 atm) with an integration step of 2 fs and saving frequency of 2 ps. After an equilibration and heating phases the production runs were started using different seeds for the initial distribution of velocities.

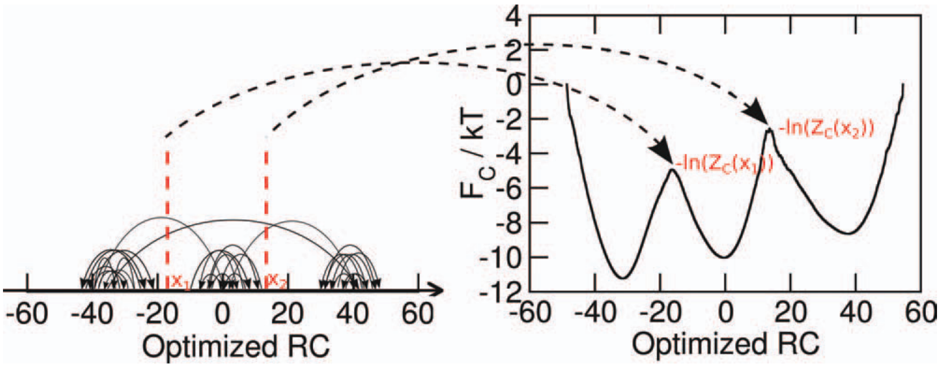


Figure 1

Schematic illustration of cFEP. By computing its optimized RC value (real number), every MD snapshot can be uniquely placed on the line representing the real numbers. The transitions are given naturally from snapshot i to snapshot $i + 1$ and are here illustrated by curved arrows (left). Then, for a fixed RC value, for example, x_1, the cut-based partition function $Z_C(x_1)$ is the number of transitions through this point, that is, the number of curved arrows crossing the dashed line at x_1. Computing this values for every x gives the cFEP as $F_C(x) = -kTln(Z_C(x))$ (right). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

The analysis of the MD trajectories was carried out with CHARMM[22] and the MD-analysis tool WORDOM.[28,29]

Optimization of reaction coordinate

To construct a one-dimensional free energy profile, the MD trajectory has to be further processed. Here, the total of 655,000 snapshots was binned along a reaction coordinate (RC), another common method would be grouping the snapshots with a clustering algorithm (e.g., using a tree-based algorithm[30]). The binning suggests the conventional way of projecting the free energy as $F_H(i) = -kTlnZ_H(i)$ where $Z_H(i)$ denotes the partition function which is equal to the density of bin i, that is, $Z_H(i) = \frac{\#\text{frames in bin } i}{\text{bin size}}$ and F_H is short for histogram-based free energy.

A more kinetically motivated approach that is able to preserve the free energy barriers is to compute the cut-based free energy profile (cFEP).[31–33] This approach emulates the cuts in flow-networks. In the case of conformational space sampled by MD, the nodes and links of the network are the mesostates (e.g., bins along a RC or clusters obtained by root mean square deviation [RMSD] clustering) and the sampled transitions between them, respectively.[34] The partition function of the free energy barrier separating two mesostates i and j is equal to the value of minimum cut between the mesostates in the network.[31] Practically, this means that the network is partitioned in two distinct sets I and J, which include mesostates i and j, respectively, such that the number of transitions between the two sets are minimal. This minimal number of transitions is then assigned as partition function of the transition state between mesostates i and j as it represents the statistical weight of this transition. Similarly, one can estimate in continuous terms the statistical weight for a RC and thereby the cut-based partition function $Z_C(x)$ at a point x, as the number of transitions through point x.[32] Accordingly, the computation of the cut-based free energy is performed as $F_C(x) = -kTln(Z_C(x))$ where x is the RC and $Z_C(x)$ is the cut-based partition function at the point x defined as the number of transitions that pass through the point x. In terms of bins, the following formula applies $Z_C(i,i+1) = \frac{1}{2}(\sum_{j \le i < k} n_{jk} + \sum_{j > i \ge k} n_{jk})$ where n_{ij} is the number of transitions from bin j to bin i. In terms of continuous RC, the formula becomes $Z_C(x) = \frac{1}{2} \sum_t \theta((x(t) - x)(x - x(t + \Delta t)))$ where θ is the Heaviside step function and Δt is the sampling interval. A schematic illustration of this procedure is shown in Figure 1.

The aim of the optimization of the RC is to find a coordinate that shows diffusive dynamics (at least at the barrier regions). Diffusive dynamics is characterized by the mean square displacement growing linearly with time. The following equation defines the so-called subdiffusion exponent α: $\langle \Delta x^2(t) \rangle \sim t^{2\alpha}$. For diffusive dynamics, one has $\alpha = 0.5$, whereas for subdiffusive dynamics $\alpha < 0.5$. The coordinate dependent exponent $\alpha(x)$ can be computed from the distance between two profiles $F_C(x)$ at different sampling intervals.[35] Plotting $F_C(x)$ together with $\alpha(x)$ gives insight into the dynamics along a profile.

The optimization procedure starts from a seed RC which is able to capture the configuration space sufficiently well, that is, partially distinguish between free energy basins, but is suboptimal, meaning that projection of the free energy onto this coordinate results in possible subdiffusive dynamics at barrier and other regions and in underestimating barrier heights due to overlapping of the configuration space. The seed RC chosen here is the RMSD of the C_α atoms of residues 6–97 from the first frame of the MD trajectory. As projecting to a suboptimal RC results in faster kinetics by overlapping of different regions of the configuration space, the optimal RC is defined by having the longest mean first passage times (MFPT) from one free energy basin to another. The optimal RC is then constructed by maximizing the MFPT that is computed from the trajectory using Kramer's equation in the following form

$$\begin{aligned}\langle t_{A,B}\rangle &= \int_a^b \mathrm{d}x \frac{e^{\beta F_H(x)}}{D(x)} \int_{-\infty}^{x} \mathrm{d}y e^{-\beta F_H(y)} \\ &= \frac{\Delta t}{\pi} \int_a^b \mathrm{d}x \frac{Z_H(x)}{Z_C^2(x)} \int_{-\infty}^{x} \mathrm{d}y Z_H(y),\end{aligned}$$

where $\beta = \frac{1}{kT}$, $D(x)$ is the coordinate-dependent diffusion coefficient, and a and b denote two representative coordinate values for two basins A and B, respectively.[32,35] The functional form of the RC is the "smoothed number of native contacts" $R(x, \alpha) = \sum_{i,j} \alpha_{ij}\, \theta(\Delta_{ij} - r_{ij})$ where r_{ij} is the distance between atoms i and j while α_{ij} and Δ_{ij} are the parameters to be optimized (the former is either 1 or -1 while the latter is a threshold for a formed contact $\in \{0Å, 0.5Å, \ldots, 30Å\}$) and θ is a smoothed step function, that is, $\theta(x) = \min(1,x)$ if $x > 0$ and zero otherwise.[35] The optimization is carried out iteratively by randomly picking a pair ij and selecting α_{ij} and Δ_{ij} according to the given ranges above, such that MFPT is maximized. In addition, pairs of coordinates ij that do not bring substantial improvement are removed automatically (see Ref. 35 for further details).

There are significant differences between the cFEP method and other frequently used techniques for projecting the free energy surface, for example, histogram-based profiles and principal component analysis. The main advantage of the cFEP method is that the dynamical information present in the trajectory is fully taken into account. Moreover, the cFEP with RC optimization does not require any geometrical clustering. By considering the actual transition matrix, the approach used in this work performs dimensionality reduction while preserving the dynamic information.[35] In contrast, other frequently used methods reduce the dimensionality by trying to preserve the proximity in configuration space. The vicinity in configuration space does not necessarily imply dynamical closeness. In other words, geometrically similar conformations can be separated by high free energy barriers.

Tilting angle

The tilting angle between α2 helix and β2 strand is a measure of scissor-like partial widening of the binding site groove. It is computed as the angle between two directional vectors given by β2 (from C_α of Ile25 to carbonyl C of Gly22) and α2 (from carbonyl C of Glu71 to carbonyl C of Lys78).

RESULTS

The following analysis is based on four explicit solvent MD runs for each of the apo and holo states. The total simulation time is 660 ns for apo PDZ3 and 650 ns for holo PDZ3 during which the native fold is preserved as expected at the temperature of 298 K. The residues 303–402 of PDZ3 according to the PDB file 1BE9 are hereafter termed residues 1–100 while the five C-terminal residues of the peptide ligand are referred to as Val0, Ser-1, Thr-2, Gln-3, and Lys-4 (from C-terminus to N-terminus).

Differences in the free energy surface and plasticity of apo and holo

The cFEP,[31] projected onto an optimized RC[35] (see "Materials and Methods" section), displays three free energy basins [called A, B, and C in Fig. 2(A)] within the native state. From the time series of the value of optimized RC [Fig. 2B (top)] along the simulation trajectories, it is evident that in the apo simulations all the three free energy basins are sampled, whereas in the holo runs the system never leaves basin A. Defining the ranges of optimized RC according to basin A,B, and C as -38 to -25, -8 to 6, and 25 to 45, respectively, every snapshot of the simulation can be allocated either to an individual basin or to the transition region according to its RC value. From this, we can derive the relative populations of the three basins in the apo simulations as approximately 36%, 45%, and 13% for basins A,B, and C, respectively, whereas the remaining 6% of the snapshots populate the barrier regions. The optimization procedure is started from an appropriate seed RC which should be able to capture the configuration space well enough. Although such a conventionally chosen RC often leads to subdiffusive dynamics and underestimating of barrier heights due to overlapping of the configuration space, the optimization procedure yields a RC which is optimal in the sense that at the barrier regions the projected dynamics are diffusive. Furthermore, possible configuration space overlap is removed. We used the RMSD of the C_α atoms of residues 6–97 from the first frame of the MD trajectory as a seed RC for the optimization . The resulting optimized RC is more appropriate than the RMSD as it separates the three free energy basins better [Fig. 2(B)] and the dynamics at the barriers on projection is diffusive [Fig. 2(A)]. The time series of the RMSD of the C_α atoms of residues 6–97 never exceeds 2 Å and is below 1.5 Å most of the time which shows that the folded state is preserved in all MD runs. Carrying out the cFEP analysis using the apo and holo sampling separately confirms the finding that apo PDZ3 samples three different free energy basins whereas holo PDZ3 samples only one. Projection of the free energy for the apo sampling onto the seed RC and later on onto the optimized RC reveals

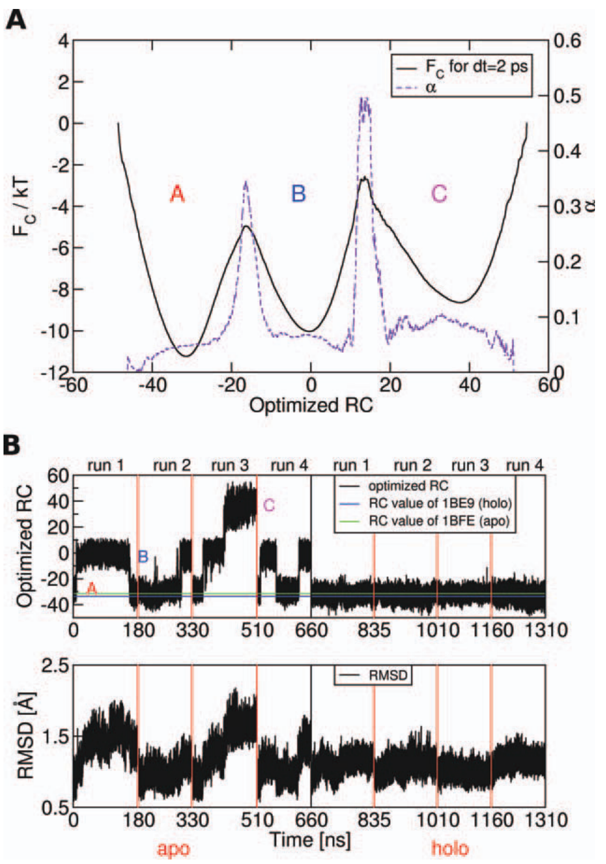


Figure 2

Only one of the three free energy basins of apo PDZ3 is populated by the PDZ3/peptide complex. (**A**) Cut-based free energy F_C along the optimized RC for the accumulated apo and holo MD runs. The three free energy basins are labeled A, B, and C. The F_C (solid line) is shown together with the subdiffusion coefficient α (dashed line). The value of the subdiffusion coefficient $\alpha = 0.5$ in the barrier regions indicates that the optimization of the RC yields diffusive dynamics[35] at the transition regions between basins in the native state. Note that the subdiffusion α < 0.5 at the bottom of the basins does not influence significantly the overall kinetics which are dominated by the barrier regions. **B**, (top) Time series of value of optimized RC along the simulation trajectories. Vertical lines indicate the beginning/end of each MD run and the black vertical line separates apo from holo runs. The time series of optimized RC shows that only the basin A is occupied in the absence of the peptide ligand while all of the three basins [labeled as in (A)] are populated in the complexed state. The value of the optimized RC for the **X**-ray structures of apo (PDB code 1BFE) and holo (PDB code 1BE9) are shown with green and blue horizontal lines, respectively. B (bottom) Time series of RMSD of the C_α atoms of residues 6–97 from the first frame. Vertical lines are shown as in the top panel. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

three free energy basins, whereas for holo PDZ3 projection of the free energy onto the seed RC results in a single free energy basin (see Supporting Information Figs. S1 and S2).

It is useful to supplement the cFEP analysis with the differences in the root mean square fluctuations (RMSF) calculated from the MD runs. The RMSF of the C_α atoms in the apo state are higher than in the complexed state (Fig. 3) which is consistent with the cFEP analysis. The RMSF differ most at the loop regions, that is, the β1–β2 and β2–β3 loops, as well as at the α-helix (α2) which lines the peptide binding groove. Overall, the RMSF calculated from the MD trajectories are in good agreement with the crystallographic temperature factors for 1BFE (apo) and 1BE9 (holo) (Fig. 3). Moreover, the X-ray structure of the apo state has higher B-factors that the one of the holo state, and the difference is larger in the same segments of the sequence as for the RMSF difference, that is, the loop regions and α2 helix.

It has to be mentioned that all of the seven X-ray structures of PDZ3 (apo: 1BFE, 1TQ3, 3I4W, and 3K82; holo: 1BE9, 1TP3, and 1TP5) are located in basin A with values of the optimized RC ranging from −23.8 for 3K82 to −34.6 for 1TP3. There are two possible reasons for the fact that according to their value of the optimized RC the four apo X-ray structures belong to basin A: it is either the most populated and/or the one favored by the crystal packing. Given such structural similarity among the available X-ray structures, the plasticity of the native state of apo PDZ3, as it emerges from the MD simulations, is rather unexpected. Thus, it is important to investigate the structural features and transitions that characterize the individual free energy basins within the native state of (apo) PDZ3, which are presented next.

Structural analysis

As indicated by the RMSF of the C_α atoms there are mainly two regions that display significant structural plasticity: the β2–β3 loop and the α2 helix which make up the binding site groove. To analyze the relative displacement of these secondary structure element, it is useful to determine the structural ensembles corresponding to each of the three basins identified by the optimized RC. The three basins can be characterized by the position of the β2–β3 loop and the relative orientation of α2 and β2 (Fig. 4). The tilting angle between α2 and β2 (for definition see Materials and Methods section) widens by about 8° and 10° in going from basin A to basin B and C, respectively (see Supporting Information Figs. S3 and S4). Similarly, NMR structures of apo and holo LARG (leukemia-associated Rho guanine nucleotide exchange factor) PDZ show that peptide binding tilts α2 and β2 apart.[10]

The distribution of selected pairs of distances are useful to further characterize the motion of the binding site [Fig. 4(C)] by complementing the observed widening of the tilting angle. The histograms of three distances between pairs of C_α atoms in α2 and β2 (Gly22-Lys78, Asn24-Ala74, Ile26-His70) indicate that the widening of the tilting angle corresponds to a scissor-like motion of α2 with respect to β2 with the pivot somewhere between the middle and the N-terminal turn of helix α2. Thus,

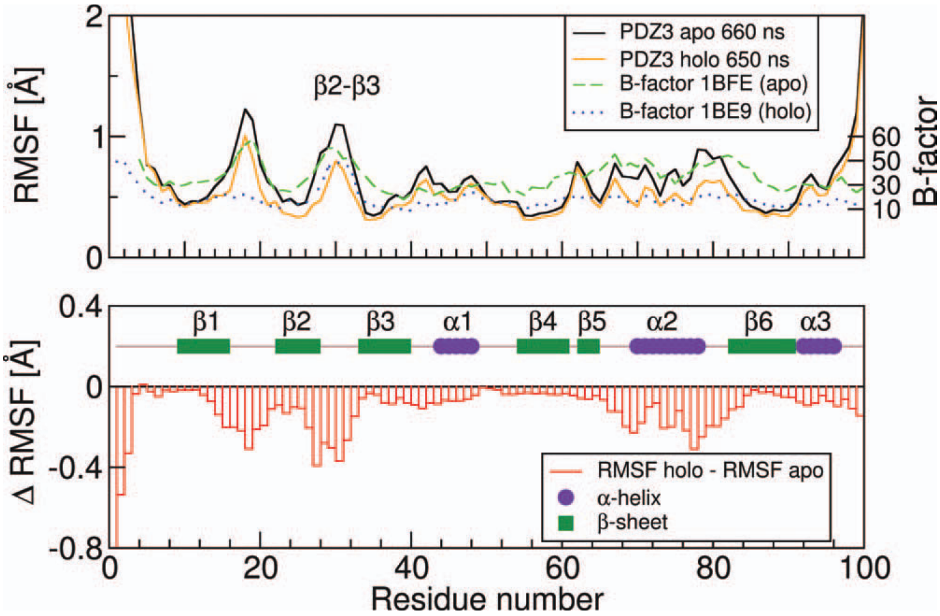


Figure 3

Differences in backbone flexibility between apo and holo PDZ3. (Top) Values of the RMSF of the C_α atoms for apo and holo PDZ3 (black and orange lines, *y*-axis on the left) and the crystallographic B-factors (blue dotted and green dashed lines, *y*-axis on the right) as a function of residue number. The RMSF values were averaged over simulation intervals of 5 ns. The first 5 ns of each MD run were discarded. (Bottom) The differences between the RMSF of holo PDZ3 and apo PDZ3 are shown by red bars. Secondary structure assignment was done with WORDOM [28,29] using the X-ray structure of holo PDZ3 (1BE9). The individual elements of secondary structure are: β1 strand, 10–15; β2 strand, 23–27; β3 strand, 34–39; β4 strand, 55–60; β5 strand, 63–64; β6 strand, 83–89; α1 helix, 44–48; α2 helix, 70–78; α3 helix, 91–97. The nomenclature used for individual α-helices and β-strands corresponds to the one introduced in Ref. 3, where β*A* to β*F* and α*A* to α*C*, are replaced by β1 to β6 and α1–α3, respectively. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the largest displacement when going from basin A to basin B and C (>2 Å) takes place at the C-terminal turn (Gly22–Lys78 distance), whereas for the N-terminal part of helix α2 (Gly22–Lys78 distance, opposite of pivot) a smaller (≈1 Å) displacement in the opposite direction is observed. This scissor-like motion, therefore, widens or narrows a hydrophobic cavity of the binding site groove located between the C-terminal part of helix α2 and the N-terminal part of strand β2.

The distance between the C_α atoms of residues 29 and 78 reflects the relative displacement of the β2–β3 loop with respect to α2. The histograms of this distance show significant differences for the three free energy basins [Figure 4(C)]. The two-peak distribution for basin B indicates that there are two main orientations of this loop in basin B which differ by about 2 Å. Since the histogram of the Gly22–Lys78 distance for basin B is single-peaked, the two peaks arise from two different positions of Glu29, that is, two different conformations of the loop. The structural overlap of representative MD snapshots of basins A, B, and C visualizes the described structural differences among basins manifested in helix α2 and the β2–β3 loop [Fig. 4(A)]. Furthermore, overlap of the available crystal structures of apo and holo PDZ3 with the representative structures of the individual basins reveals, as suggested by analysis of RC values, that apo and holo X-ray structures match best with the representative snapshot of basin A. Importantly, a tilting angle of helix α2 comparable to the one of the representative snapshot of basin B is not observed among the crystal structures [Fig. 4(D)].

The α3 helix is a structural element of PDZ3 that is additional to the usual PDZ fold, and has been previously suggested to modulate binding by a dynamic mechanism.[14] Intriguingly, α3 includes four charged residues (Lys91, Glu93, Glu94, and Arg97) that form a complex network of salt bridges. It is, therefore, useful to relate individual basins of the cFEP to the presence/absence of salt bridges that “connect” the helix α3 to the rest of the fold. Interestingly, there are differences between apo and holo state in this network of salt bridges. In particular, Arg97 is involved in a salt bridge with Glu29 or Glu32 in the apo state and almost only with the latter in the complexed state (Fig. 5). Moreover, the salt bridge between Arg7 and Glu93 is almost always formed in the holo state while often disrupted in apo PDZ3. On the other hand, the Glu94-Arg97 salt bridge and the Lys91-Glu94 water-bridged interaction, both within helix α3, are almost

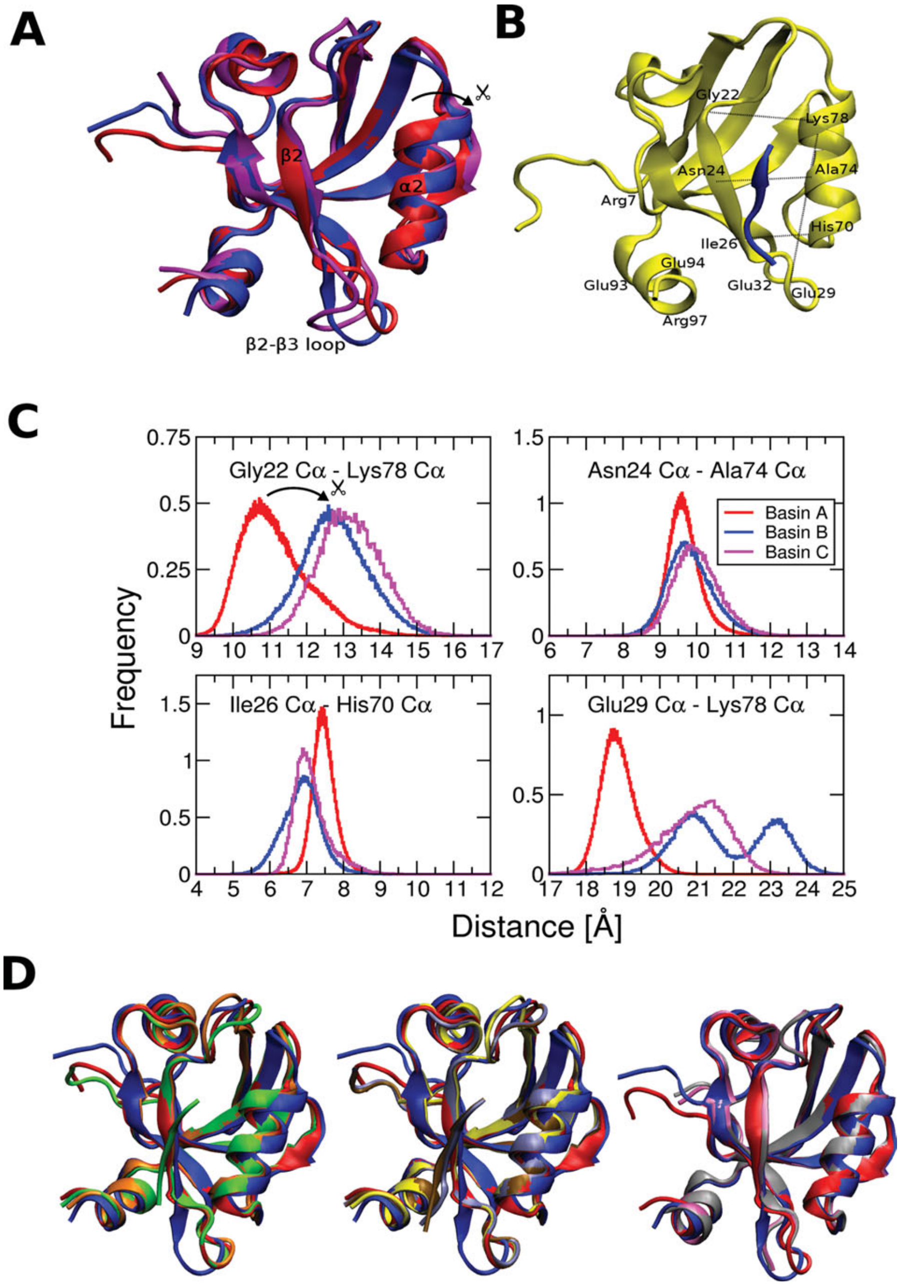


Figure 4

Structural characteristics of free energy basins of apo PDZ3. (**A**) Structural overlap of the representative structures of basins A (red), B (blue), and C (magenta). The representative structure is the MD snapshot within the smallest RMSD of the C_α atoms of residues 6–97 from the average structure of the basin. The representatives of basins A, B, and C differ from their respective average structure by RMSD values of 0.36 Å, 0.56 Å, and 0.47 Å, respectively (for distribution of RMSD values within basins see Supporting Information Fig. S5). (**B**) Snapshot of holo MD simulations showing the binding site groove with the bound peptide and indicating the locations of key residues used for the analysis of the tilting of helix α2 and the network of salt bridges. (**C**) Histograms of distances between pairs of C_α atoms in the binding site. The MD snapshots from all runs with value of the optimized RC in the range −38 to −25, −8 to 6, and 25 to 45 were used to calculate the histograms of basins A, B, and C, respectively. The colors used for the individual histograms are consistent with those used for the structures shown in the top, left. (**D**) Overlap of X-ray structures and MD snapshots representative of basins A and B. In all three figures, the representative snapshot of basin A is shown in red and the one of basin B in blue. The following X-ray structures are shown: (Left) 1BFE (apo) in orange and 1BE9 (holo) in green; (Middle) 1TQ3 (apo) in yellow, 1TP3 (holo) in ochre, and 1TP5 (holo) in iceblue; (Right) 3I4W (apo) in mauve, and 3K82 (apo) in silver. All structures were overlapped using the C_α atoms of residues 6–97.

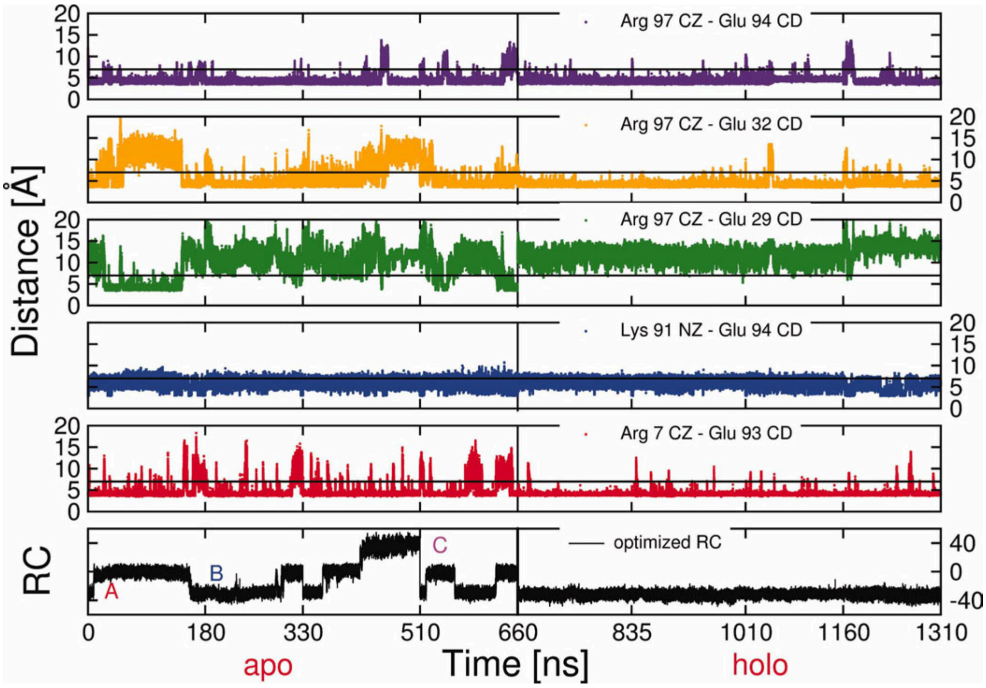


Figure 5

Salt bridges involving residues in the α3 helix. Time series of salt bridge distances between residues in helix α3 and the rest of the domain (color lines) are shown together with the time series of the optimized RC (black line, bottom panel). The black vertical lines at 660 ns indicate the concatenation point of apo and holo trajectories and the tick marks along the time axis indicate the individual apo and holo MD runs. The black horizontal lines at distance = 7 Å is drawn to discriminate between formation and rupture of salt bridges. This rather large threshold takes also into account single water-bridged polar interactions. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

always present and their behavior is indistinguishable between ligand-free and complexed state. It is clear from Figure 5 that the formation/rupture of these salt bridges is sometimes but not always concomitant with changes in the optimized RC, that is, transitions between free energy basins.

As noted above, one of the main structural differences between the three basins is the position and conformation of the β2–β3 loop. In this context, the salt bridge Arg97-Glu29 plays an important role, as its formation or rupture has a direct impact on the conformation of this loop. While this salt bridge is almost never formed in the complexed state, it spontaneously forms several times during the apo simulations, particularly in basin B (Fig. 5), thereby directly modifying the conformation of the β2–β3 loop. This simulation result is interesting because experimental studies have provided evidence of the involvement of the β2–β3 loop in binding. Using 2D and 3D heteronuclear NMR spectroscopy, Kozlov et al.[5] have reported that this loop is directly involved in the interactions between the peptide ligand and human PDZ2. In addition, Liu et al.[10] have observed by NMR experiments a high conformational flexibility for the β2–β3 loop of LARG PDZ in the ligand-free state in contrast to relative rigidity in the complexed state, and have speculated that conformational plasticity might facilitate binding of different peptide ligands. Our MD simulations indicate that the chain of hydrogen bonds involving the carbonyl oxygen of Gly27, the imidazole ring of His70, and the hydroxyl oxygen of Thr-2 in the ligand, originally observed in the X-ray structure of the PDZ3/peptide complex,[3] restricts the β2–β3 loop's flexibility in the complexed state and keeps it fixed in a single conformation. This chain of hydrogen bonds is present throughout the holo MD simulations (Fig. 6). Furthermore, it has been shown experimentally, that the mutation of Gly27 into an alanine decreases the peptide affinity by almost 100-fold, which is most probably due to the bulkier side chain of alanine preventing the β2–β3 loop from adopting the right conformation to form the hydrogen bond with His70.[36] Whether the flexibility of the β2–β3 loop facilitates or hinders binding by providing conformers to which the peptide can not bind is not evident from the simulations.

Side chain flexibility

Some but not all residues of the PDZ3 binding site show enhanced side chain flexibility in the ligand-free state compared to the complexed state (Fig. 7). The con-

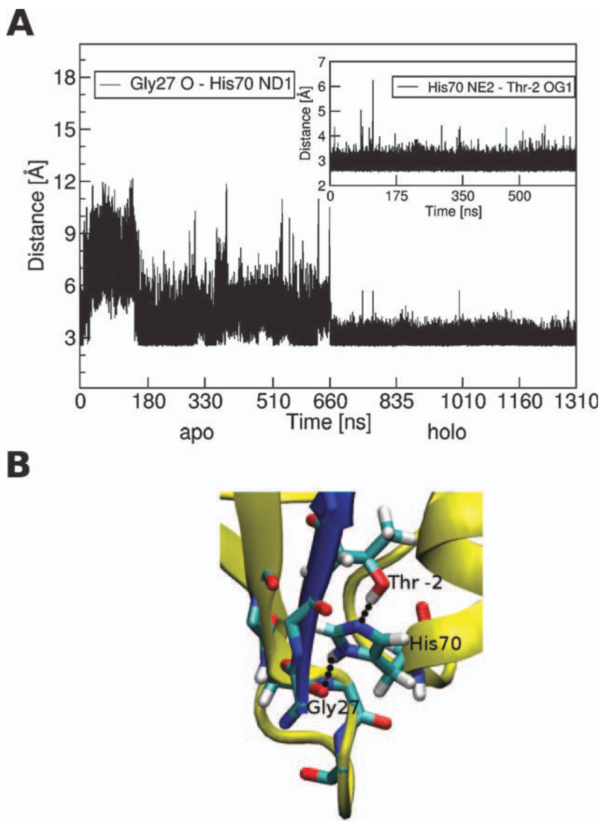


Figure 6

His70 hydrogen bonds. (A) Time series of distances between the atoms involved in the two hydrogen bonds formed by His70 in the crystal structure of the complexed state.[3] The inset shows the intermolecular hydrogen bond. (**B**) β2–β3 loop and part of the peptide binding pocket taken from an arbitrary snapshot of holo MD simulations. The backbone of PDZ3 (yellow) and peptide ligand (blue) are shown in a ribbon representation while residues involved in hydrogen bonds with His70 are shown by sticks colored by atom type. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

trary is not observed. Three different χ_1 values are accessible to His70 in the ligand-free state while only the trans rotameric state is populated in the complex because, as aforementioned, the His70 side chain is directly involved in peptide binding (Fig. 6). Furthermore, the side chains of Asn24 and Lys78, which are partially exposed to solvent, show a decreased flexibility in the complexed state. While Asn24 forms a hydrogen bond with Gln-3 of the peptide ligand in the crystal structure of the complex,[3] for Lys78 no such interactions have been suggested. Interestingly, the MD simulations reveal that the side chain amino group of Lys78 can act as donor in a hydrogen bond with the carbonyl oxygen of Ser-1 (see Supporting Information Fig. S6 for time series of PDZ3/peptide hydrogen bonds). The χ_1-angle flexibility of Leu77 decreases upon peptide binding. Leu77 is part of the hydrophobic cavity of the binding site, as is Phe23, that is filled by the side chain of Val0 of the peptide ligand. The crystal structure of the complexed state displays a hydrogen bond between the carbonyl oxygen of Leu77 and a guanidinium nitrogen of Arg16. However, the latter interaction is equally probable for ligand-free and complexed PDZ3 (see Supporting Information Fig. S6). Therefore, the decreased flexibility of Leu77 in holo state is most likely due to the further burial of Leu77 by the side chain of Val0. There are also residues in β2 strand and α2 helix that show similar flexibility in the holo and apo state. As an example, the χ_1 of Phe23 populates a single rotameric state because its phenyl ring is tightly packed in the hydrophobic part of the binding pocket.

DISCUSSION

We have investigated the differences between the complexed and peptide-free states of the third PDZ domain of the synaptic protein PSD-95 by multiple explicit solvent MD simulations and analysis of free energy surface. Two main results emerge from our simulation study. First, the free energy landscape of the apo native state is more complex than the one of the peptide-bound state. For apo PDZ3 there are three main basins separated by barriers of about 3–4 kcal/mol. The presence of multiple free energy basins in the native state of apo PDZ3 is consistent with the folding intermediate observed under native conditions by native-state hydrogen exchange experiments.[37] In contrast, the holo state consists of a single basin which implies that there is an entropic penalty in binding the peptide ligand, in agreement with the conformational entropy mechanism and dynamic allostery unveiled by NMR experiments of wild type and a C-terminal truncated form of PDZ3.[14] Interestingly, the additional C-terminal helix α3 (residues 91–97), which has been shown to modify side chain dynamics, plays also a structural role in binding by influencing the conformation of the β2–β3 loop (residues 28–33) via a network of salt bridges involving Lys91, Glu94, and Arg97 in the former, and Glu29 and Glu32 in the latter.

Second, the cFEP analysis shows that the holo state samples only one conformation, which is also one of the three conformations sampled by the apo state. Thus, the peptide ligand binds to PDZ3 by selection of one of the three conformations sampled by the apo state. As the present simulation study does not sample the binding process explicitly, binding to one of the two apo-only conformations followed by spatial rearrangement (to the conformation of basin A) cannot be conclusively excluded. Nevertheless, experimental evidence supports a one-step binding mechanism without rate-limiting conformational change.[7,8] Therefore, the theoretical framework of the conformational selection binding mechanism is the one most in line with our observations. The different types of analysis of the MD trajectories provide con-

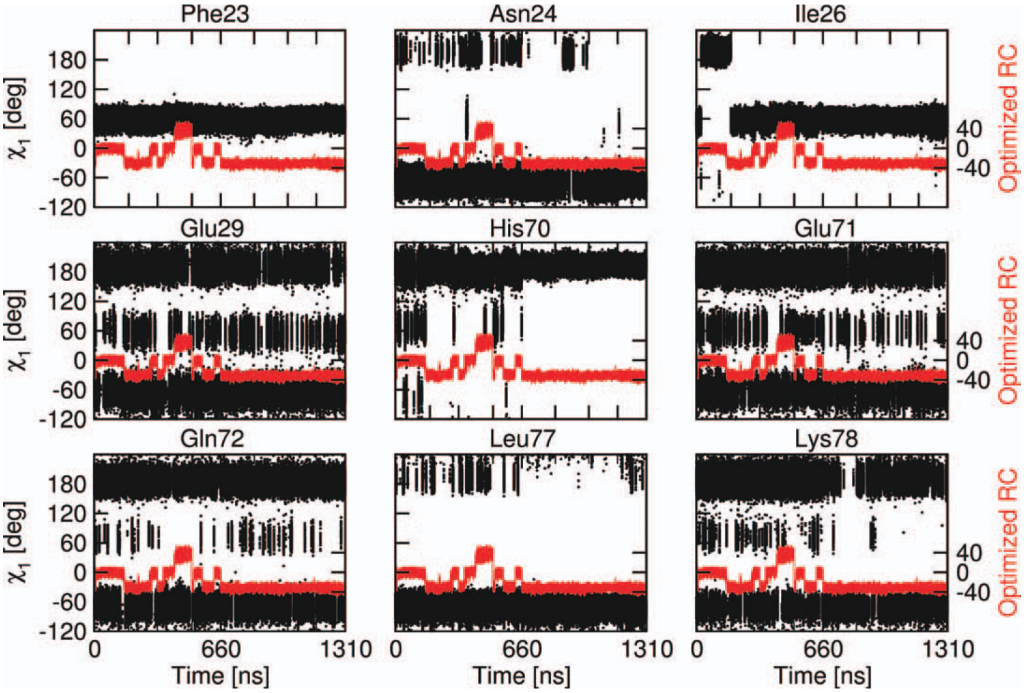


Figure 7

Flexibility of side chains. Time series of χ_1-angles of binding site residues (black, *y*-axis on the left) together with the time series of the optimized RC (red, *y*-axis on the right). The tick mark at time = 660 ns indicates the separation of apo and holo trajectories, additional tick marks on the *x*-axis indicate the independent apo and holo MD runs. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

sistent evidence that the native conformation of PDZ3 is more heterogeneous in the apo state than the holo state. Moreover, the increased plasticity in the apo state is not restricted to side chain fluctuations, as suggested by NMR spectroscopy data,[14] but involves also the backbone and in particular the relative orientation of the regular elements of secondary structure lining up the binding site (β2 strand and α2 helix) as well as the β2–β3 loop. The scissor-like motion of helix α2 and the thereby induced partial widening of the binding site groove takes place on a time scale of 10–100 ns and is reversible. In contrast, the single conformer of the peptide-bound state of PDZ3 is the one with the smallest tilting angle of the α2 helix with respect to the β2 strand.

Importantly, conformational selection as peptide-binding mechanism for a PDZ domain is a novel suggestion. It remains to be investigated if it applies also to other PDZ domains. Up to now, experimental data on binding kinetics is available for some but not all PDZ domains, for example for the murine form of PDZ2 [8] and the second PDZ domain of SAP97.[9] Nevertheless, in several cases where such data is not available, hypotheses for the binding mechanisms were made by analysis of crystal structures and/or NMR spectroscopy data of PDZ domains. The present simulation study shows that information on the free-energy landscape obtained by MD, even if restricted to the native basin, can significantly alter the picture of the binding mechanism suggested on the basis of X-ray crystallography and/or NMR spectroscopy data.

It is interesting to compare the behavior of PDZ3 with other PDZ domains. The present MD simulations reveal that the conformation accessible to the complexed form of PDZ3 is the one with the smallest aperture of the hydrophobic cavity of the binding site groove, whereas the ensembles of NMR structures of apo and holo murine PDZ2 of PTP-BL[8] show the exact contrary, that is, an opening of the binding site on peptide binding by PDZ2. Note also that the accessibility of different tilting angles/orientation of the α2 helix with respect to the β2 strand has also been shown experimentally for other PDZ domains as LARG PDZ[10] and the PDZ domain of Par-6.[38] However, the way this tilt is modified on peptide binding, and thereby widens or narrows the binding site groove, might depend not only on the specific PDZ domain itself but also on the bulkiness of the side chains of the peptide ligand.

ACKNOWLEDGMENTS

The authors thank Sergei Krivov for very interesting discussions. The simulations were carried out on the Schrödinger compute cluster of the University of Zurich.

REFERENCES

1. Kim E, Sheng M. PDZ domain proteins of synapses. Nat Rev Neurosci 2004;5:771–781.
2. Fan J-S, Zhang M. Signaling complex organization by PDZ domain proteins. Neurosignals 2002;11:315–321.
3. Doyle DA, Lee A, Lewis J, Kim E, Sheng M, MacKinnon R. Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. Cell 1996;85:1067–1076.
4. Walma T, Spronk CA, Tessari M, Aelen J, Schepens J, Hendriks W, Vuister GW. Structure, dynamics and binding characteristics of the second PDZ domain of PTP-BL. J Mol Biol 2002;316:1101–1110.
5. Kozlov G, Banville D, Gehring K, Ekiel I. Solution structure of the PDZ2 domain from cytosolic human phosphatase hPTP1E complexed with a peptide reveals contribution of the β2-β3 loop to PDZ domain-ligand interactions. J Mol Biol 2002;320:813–820.
6. Zhang J, Sapienza PJ, Ke H, Chang A, Hengel SR, Wang H, Phillips GN, Lee AL. Crystallographic and nuclear magnetic resonance evaluation of the impact of peptide binding to the second PDZ domain of protein tyrosine phosphatase 1e. Biochemistry 2010;49:9280–9291.
7. Gianni S, Engström A, Larsson M, Calosci N, Malatesta F, Eklund L, Ngang CC, Travaglini-Allocatelli C, Jemth P. The kinetics of PDZ domain-ligand interactions and implications for the binding mechanism. J Biol Chem 2005;280:34805–34812.
8. Gianni S, Walma T, Arcovito A, Calosci N, Belli A, Engström A, Travaglini-Allocatelli C, Brunori M, Jemth P, Vuister GW. Demonstration of long-range interactions in a PDZ domain by NMR, kinetics, and protein engineering. Structure 2006;14:1801–1809.
9. Chi CN, Bach A, Engström A, Wang H, Strømgaard K, Gianni S, Jemth P. A sequential binding mechanism in a PDZ domain. Biochemistry 2009;48:7089–7097.
10. Liu J, Zhang J, Yang Y, Huang H, Shen W, Hu Q, Wang X, Wu J, Shi Y. Conformational change upon ligand binding and dynamics of the PDZ domain from leukemia-associated Rho guanine nucleotide exchange factor. Protein Sci 2008;17:1003–1014.
11. van den Berk LCJ, Landi E, Walma T, Vuister GW, Dente L, Hendriks JAJ. An allosteric intramolecular PDZ-PDZ interaction modulates PTP-BL PDZ2 binding specificity. Biochemistry 2007;46:13629–13637.
12. Zhang M. Scaffold proteins as dynamic switches. Nat Chem Biol 2007;3:756–757.
13. Sohn J, Grant RA, Sauer RT. Allostery is an intrinsic property of the protease domain of DegS: implications for enzyme function and evolution. J Biol Chem 2010;285:34039–34047.
14. Petit CM, Zhang J, Sapienza PJ, Fuentes EJ, Lee AL. Hidden dynamic allostery in a PDZ domain. Proc Natl Acad Sci USA 2009;106:18249–18254.
15. Gerek ZN, Ozkan SB. Change in allosteric network affects binding affinities of PDZ domains: analysis through perturbation response scanning. PLoS Comput Biol 2011;7:e1002154.
16. Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. Science 1999;286:295–299.
17. Kong Y, Karplus M. Signaling pathways of PDZ2 domain: a molecular dynamics interaction correlation analysis. Proteins 2009;74:145–154.
18. Csermely P, Palotai R, Nussinov R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. Trends Biochem Sci 2010;35:539–546.
19. Silva D-A, Bowman GR, Sosa-Peinado A, Huang X. A role for both conformational selection and induced fit in ligand binding by the LAO protein. PLoS Comput Biol 2011;7:e1002054.
20. Ekonomiuk D, Caflisch A. Activation of the West Nile virus NS3 protease: molecular dynamics evidence for a conformational selection mechanism. Protein Sci 2009;18:1003–1011.
21. Staneva I, Wallin S. Binding free energy landscape of domain-peptide interactions. PLoS Comput Biol 2011;7:e1002131.
22. Brooks BR, Brooks CL, MacKerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: the biomolecular simulation program. J Comput Chem 2009;30:1545–1614.
23. van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: fast, flexible, and free. J Comput Chem 2005;26:1701–1718.
24. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Jr., Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiœrkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.
25. Jorgensen WL, Chandrasekhar J, Madura J, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
26. Darden T, York D, Pedersen LG. Particle mesh Ewald: an nlog(n) method for Ewald sums in large systems. J Chem Phys 1993;98:10089.
27. Berendsen HJC, Postma JPM, van Gunsteren WF, Dinola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.
28. Seeber M, Cecchini M, Rao F, Settanni G, Caflisch A. Wordom: a program for efficient analysis of molecular dynamics simulations. Bioinformatics 2007;23:2625–2627.
29. Seeber M, Felline A, Raimondi F, Muff S, Friedman R, Rao F, Caflisch A, Fanelli F. Wordom: a user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. J Comput Chem 2011;32:1183–1194.
30. Vitalis A, Caflisch A. Efficient construction of mesostate networks from molecular dynamics trajectories. J Chem Theor Comput 2012;8:1108–1120.
31. Krivov SV, Karplus M. One-dimensional free-energy profiles of complex systems: progress variables that preserve the barriers. J Phys Chem B 2006;110:12689–12698.
32. Krivov SV, Karplus M. Diffusive reaction dynamics on invariant free energy profiles. Proc Natl Acad Sci USA 2008;105:13841–13846.
33. Krivov SV, Muff S, Caflisch A, Karplus M. One-dimensional barrier preserving free-energy projections of a beta-sheet miniprotein: new insights into the folding process. J Phys Chem B 2008;112:8701–8714.
34. Rao F, Caflisch A. The protein folding network. J Mol Biol 2004;342:299–306.
35. Krivov SV. Is protein folding sub-diffusive? PLoS Comput Biol 2010;6:e1000921.
36. Chi CN, Elfström L, Shi Y, Snäll T, Engström A, Jemth P. Reassessing a sparse energetic network within a single protein domain. Proc Natl Acad Sci USA 2008;105:4679–4684.
37. Feng H, Vu N-D, Bai Y. Detection of a hidden folding intermediate of the third domain of PDZ. J Mol Biol 2005;346:345–353.
38. Peterson FC, Penkert RR, Volkman Brian F, Prehoda KE. Cdc42 regulates the Par-6 PDZ domain through an allosteric CRIB-PDZ transition. Mol Cell 2004;13:665–676.

Supplementary Material for: Peptide binding to the PDZ3 domain by conformational selection

S. Steiner[a], A. Caflisch*,[a]

[a]*Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland*
**corresponding author, tel: +41 44 635 55 21, fax: +41 44 635 68 62, e-mail: caflisch@bioc.uzh.ch*

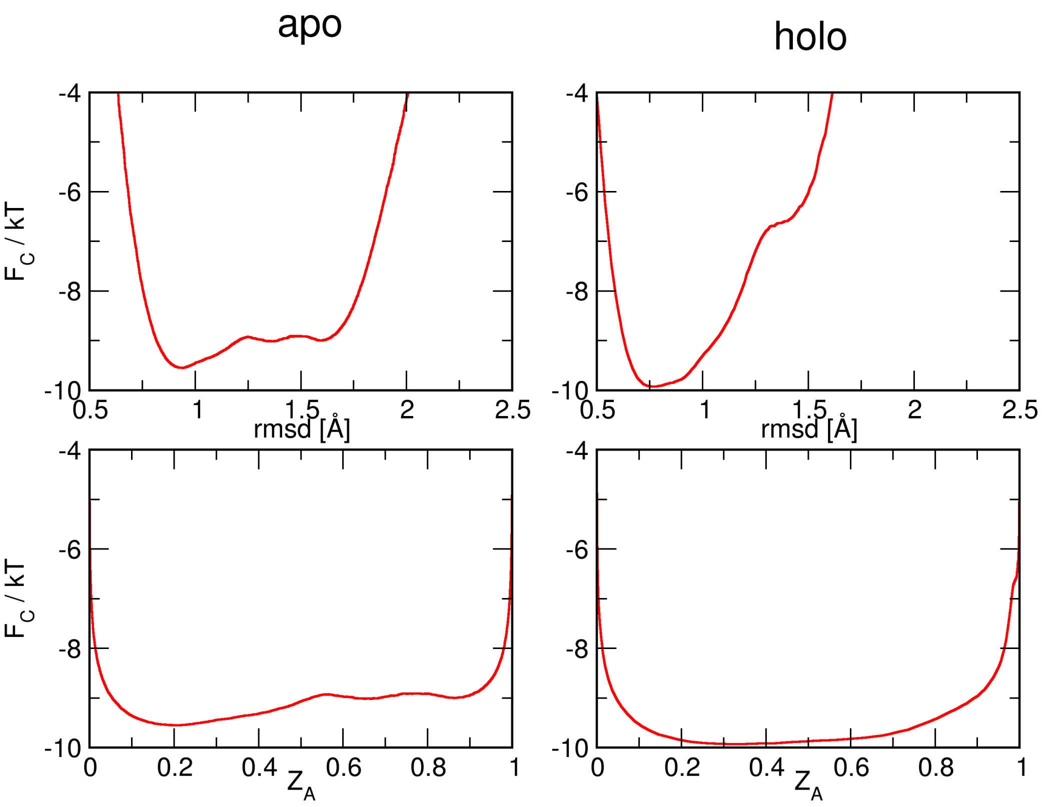


Figure S1: **Cut-based free energy profiles for apo and holo PDZ3.** The root mean square deviation (RMSD) of the C_α atomss of residues 6 to 97 from the first frame computed along the apo or holo trajectory was used as reaction coordinate (RC). The top plots show the profiles projected onto the RMSD while the bottom plots show the projection onto the relative partition function Z_A as introduced by Krivov[31]. Apo PDZ3 shows three distinguishable subbasins while the profile of holo PDZ3 consists of a single free energy basin. The lack of barriers for holo PDZ3 hinders the optimization of RC. For apo PDZ3, the cut-based free energy profile projected onto the optimized RC is shown in S2.

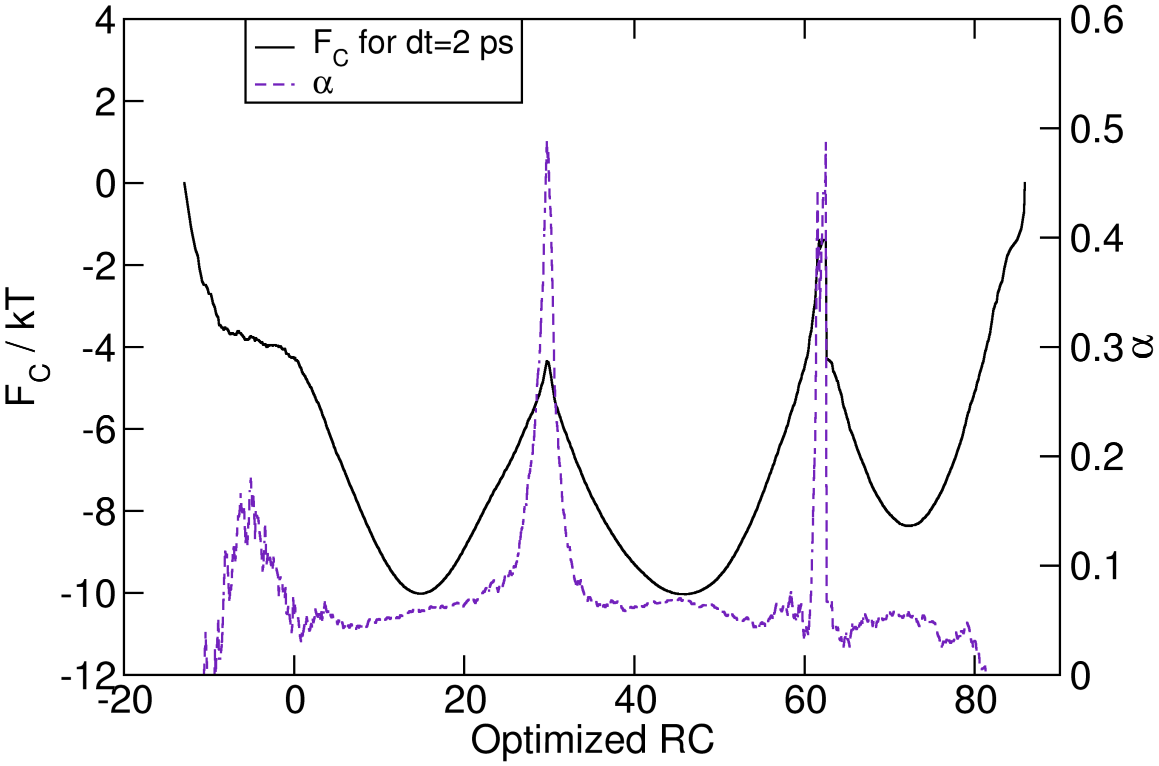


Figure S2: **Free energy profile of apo PDZ3 along optimized RC.** The cut-based free energy F_C along the optimized RC (seed coordinate was RMSD) for sampling interval $dt = 2$ ps (solid line) together with the sub-diffusion coefficient α (dashed line). Three free energy subbasins are present and they correspond to the subbasins observed when constructing the cFEP using RMSD as RC.

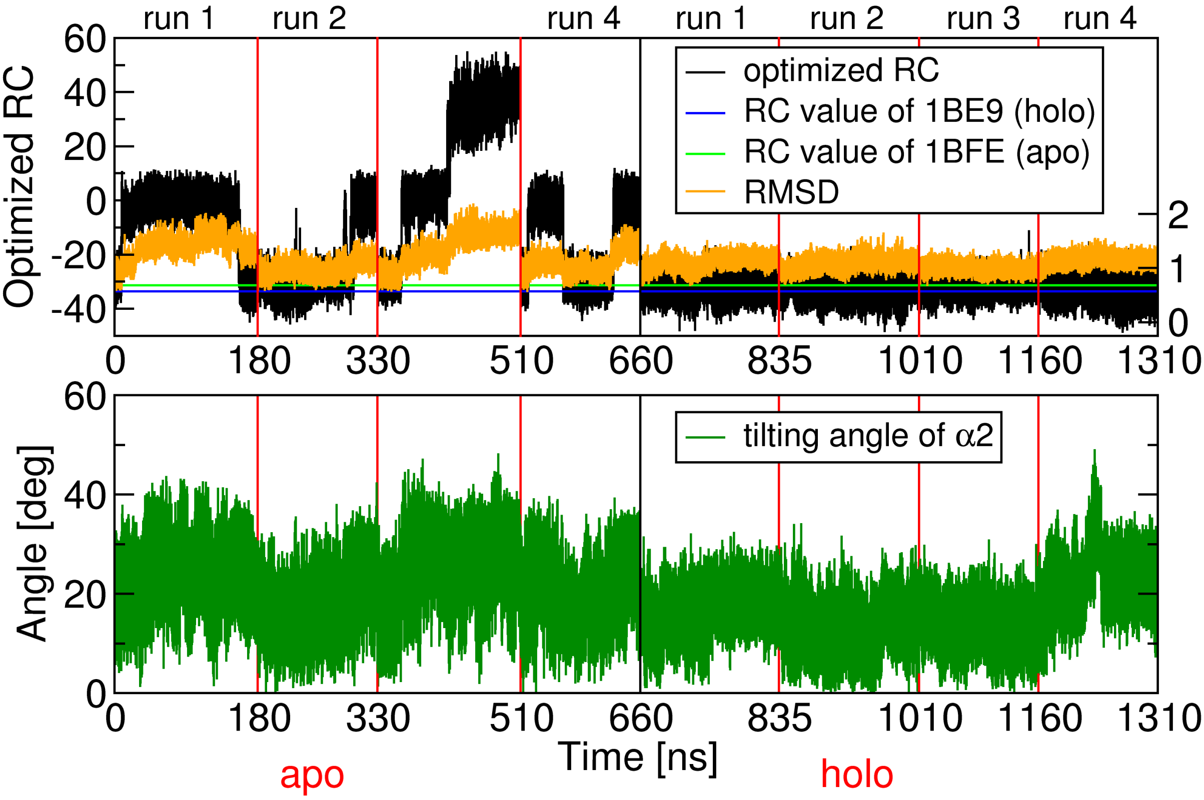


Figure S3: **Tilting angle and optimized RC.** Time series of tilting angle between $\alpha 2$ and $\beta 2$ (green, bottom panel) along the total trajectory together with the time series of the RMSD (orange, top panel, y-axis on the right) and the optimized RC (black, top panel, y-axis on the left). The tilting angle between $\alpha 2$ and $\beta 2$ was computed as the angle between two directional vectors given by $\beta 2$ (from C_α of Ile25 to carbonyl C of Gly22) and $\alpha 2$ (from carbonyl C of Glu71 to carbonyl C of Lys78). Vertical lines indicate the beginning/end of each MD run and the black vertical line separates apo from holo runs. The values of the optimized RC of the X-ray structures of apo (pdb code 1BFE) and holo (pdb code 1BE9)[3] are shown with green and blue horizontal lines, respectively (top panel).

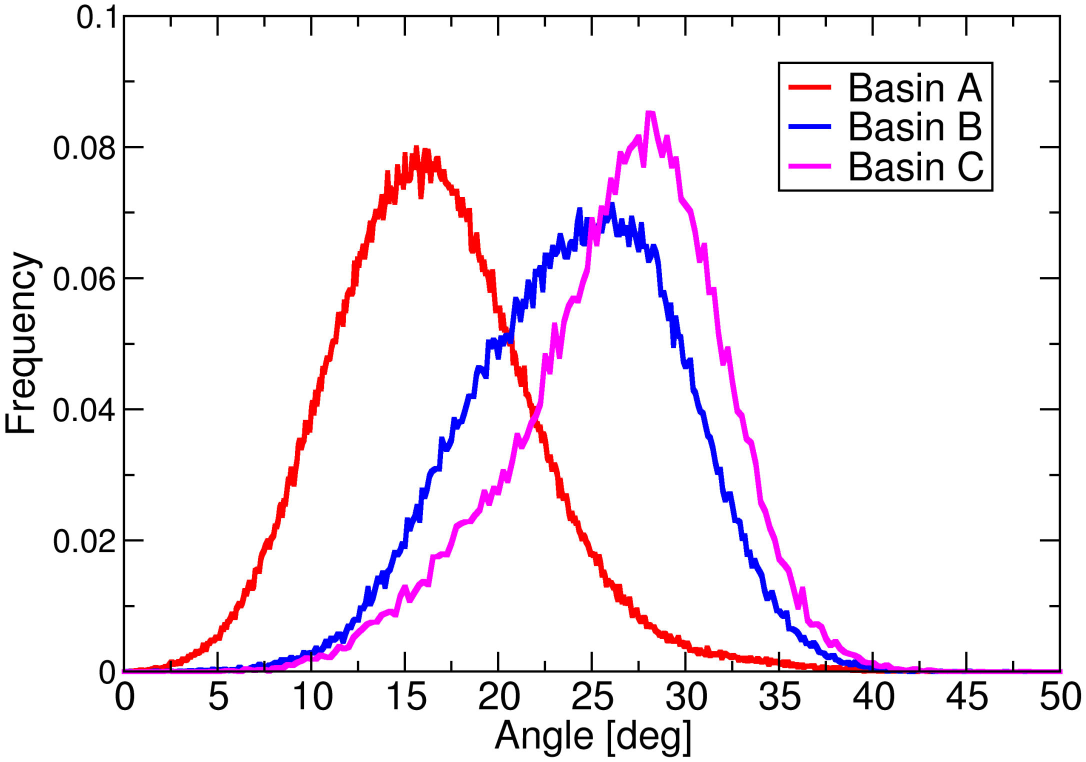


Figure S4: **Basin specific histograms for the tilting angle.** Histograms of $\alpha2$ tilting angle for subbasins A, B and C. For subbasin A the average angle is 16 with a standard deviation of 5, for subbasin B it is average $= 24 \pm 6$, and for subbasin C average $= 26 \pm 5$.

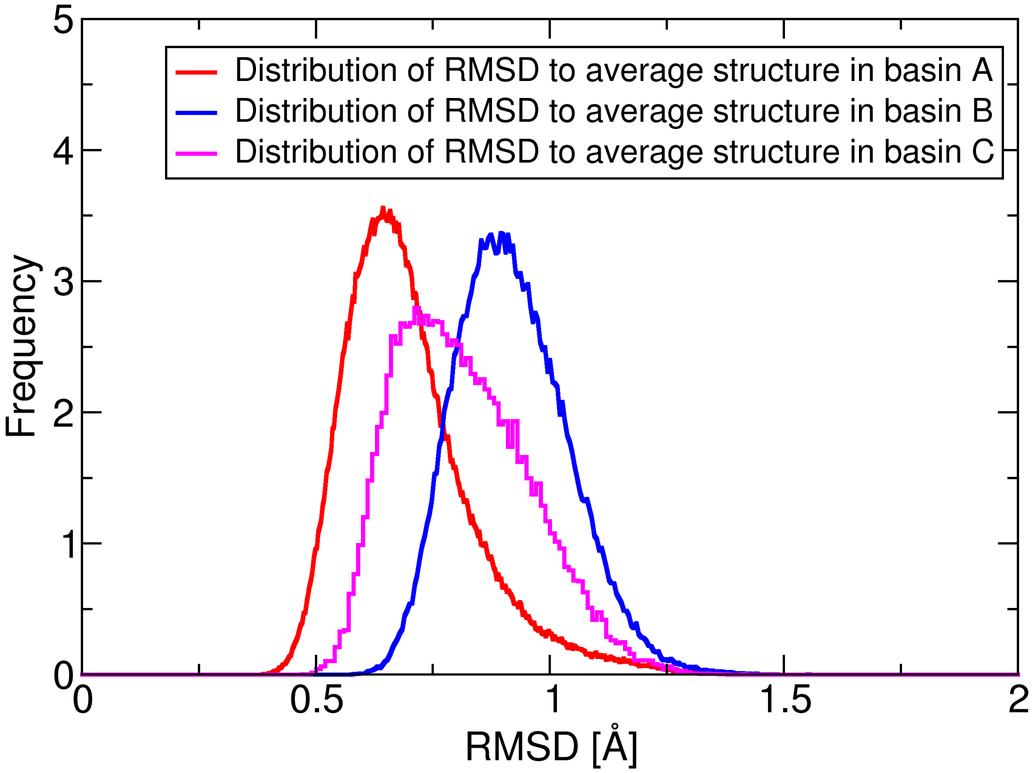


Figure S5: **Basin specific histograms of RMSD values from average structure.** For each basin the average structure was computed in order to find a representative snapshot. Here histograms of RMSD values from average structure within the respective basins are plotted. The RMSD was computed for the C_α atoms of residues 6 to 97.

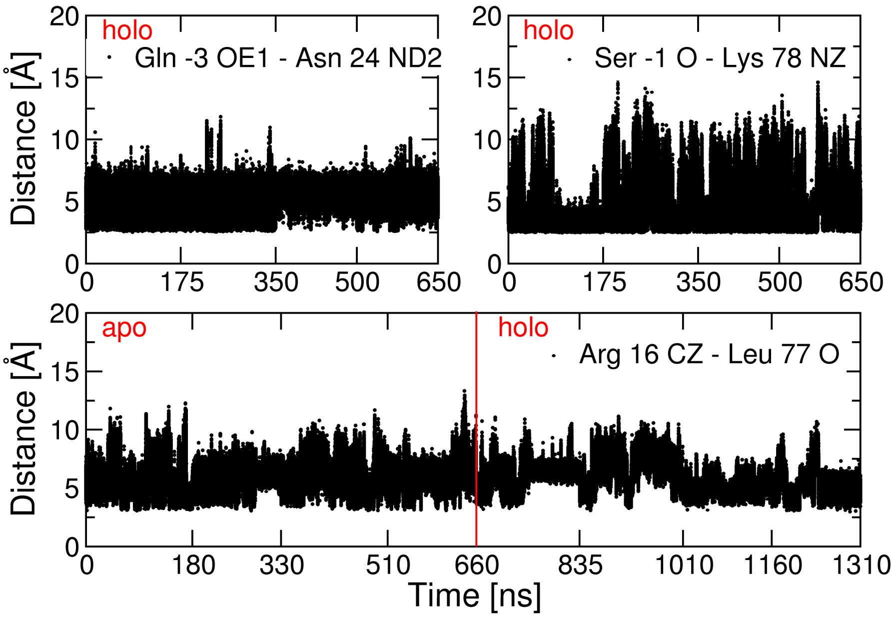


Figure S6: **Further hydrogen bonds involved in binding.** Time series of distances between atoms involved in different hydrogen bonds found in the crystal structure of complexed PDZ3[3]. For Arg16 the central carbon atom of the guanidinium group was used in the distance calculation to account for the symmetry. The red line in the bottom plot marks the concatenation point of apo and holo trajectories while the tick marks on all the three x-axes represent the concatenation of individual, independent MD runs.

Chapter 3

Peptide unbinding from PDZ2

3.1 Introduction

An introduction to PDZ domains in general is given in Section 1.1.1. As noted there, PDZ domains bind their ligand proteins by binding their C-termini in an extended conformation to an elongated pocket between the $\beta 2$ strand and the $\alpha 2$ helix, thereby augmenting the antiparallel $\beta 2$-$\beta 3$ hairpin into a three-stranded β-sheet (see Figure 1.2 and 3.1). Different PDZ domains are able to recognize different peptide-sequences via different interaction providing residues in the binding site. They thus show a varied selectivity for up to seven C-terminal ligand residues [1]. In fact, a coarse classification of PDZ domains by preference of C-terminal ligand sequences has shown three canonical classes of PDZ domains: class I prefers ligand motifs of $[X-(S/T)-X-(V/I)_{COOH}]$, class II $[X-\Phi-X-\Phi_{COOH}]$ and class III $[X-(D/E)-X-\Phi_{COOH}]$, where X is an arbitrary amino acid and Φ is a hydrophobic residue (see [2, 3]). Thus, a hydrophobic residue at the C-terminus is preferred by all the PDZ domains. This is due to the fact that they usually have a hydrophobic cavity in the binding site, where the hydrophobic side chain of the most C-terminal residue can dip in. The interaction of the C-terminal residue is considered to be the most relevant for binding, due to its hydrophobic interaction with the cavity and the interactions of the carboxylate group (see Figure 3.1). Furthermore, two structural studies indicate a (meta)stable binding mode where the peptide ligand inserts only the C-terminal residue, while the other residues remain perpendicular to the protein binding site [1, 4]. These two studies comprise NMR and crystallographic data, in fact the “perpendicular” binding pose has been found with NMR for PDZ1 of $X11\alpha$ [4] and in crystallographic structures for PDZ domains of GRASP, PDLIM2, PDLIM5 and PDLIM7 [1]. In line with these findings, a paper of Per Jemth and coworkers [5] studies the stability and relevance for binding of the individual intermolecular backbone hydrogen bonds (Figure 3.1) using amide-to-ester mutations. According to this study, removal of single backbone hydrogen bonds leads to a severe loss of affinity and formation of some of the backbone hydrogen bonds is partially rate-limiting.

Here, we present an excerpt of ongoing research and compare it with the previously mentioned studies. We have simulated the second PDZ domain (PDZ2) of human tyrosine phosphatase hPTP1E (PDZ2) complexed with a 6 residue C-terminal peptide derived from the guanine nucleotide exchange factor RA-GEF-2 (PDB code 3LNY), and we have found two complete and one partial unbinding event in four independent simulations. The analysis of timeseries of intermolecular distances shows that N-terminal backbone hydrogen bonds are disrupted before C-terminal ones. Additionally, we observe a metastable state after partial unbinding, where only the C-terminal residue is bound. This conformation can be stabilized either by the native interactions formed by the ligand’s C-terminus (i.e., intermolecular backbone hydrogen bonds and interactions of the carboxylate group, see Figure 3.1) or just by non-native interactions of the carboxylate group.

3.2 Results and Discussion

Four independet molecular dynamics (MD) runs each of length 1.3 μs of PDZ2 in complex with RA-GEF-2 peptide were performed. The peptide has the following sequence from C- to N-terminus: Val0, Ala-1, Ser-2, Val-3, Gln-4, Glu-5. The residues are numbered decreasingly starting with 0 at the C-terminus.

In two of these runs we observe unbinding, in one we observe a partial unbinding

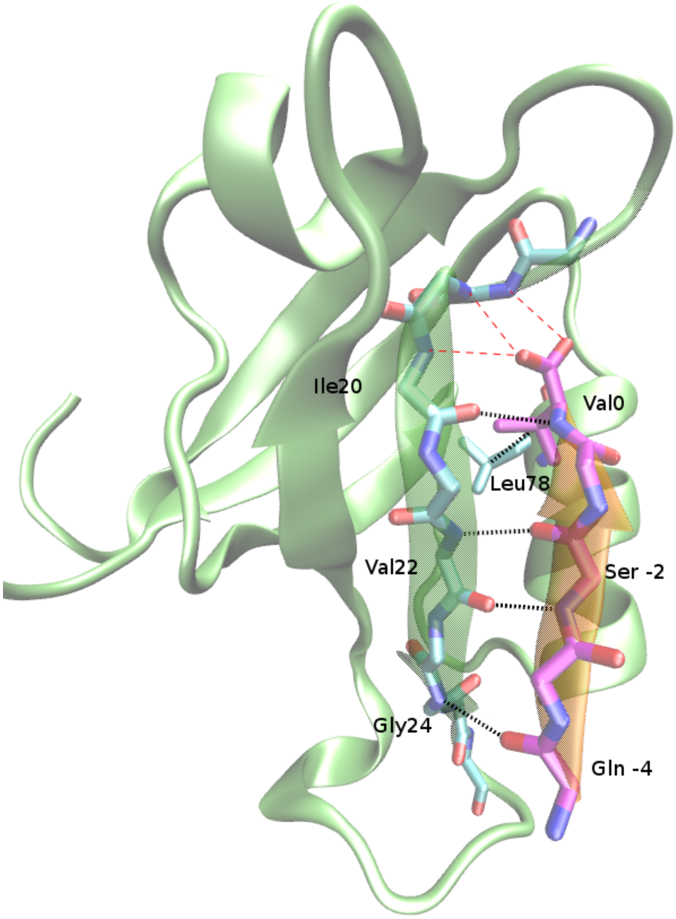


Figure 3.1: **Intermolecular backbone hydrogen bonds.** The intermolecular backbone hydrogen bonds used for analysis are shown as black dashed lines. Furthermore, the distance between C_β of Val0 and C_γ of Leu78 is shown as black dashed line, which was employed to determine the detachment of the various hydrophobic groups in the binding site cavity. The interactions formed by the carboxylate group of Val0 are indicated by red dashed lines. (Residue Glu-6 of the ligand is not shown.)

event. Figure 3.2 shows the time series of the distance of the geometric center of the ligand and the geometric center of the binding site. In the bound state this distance fluctuates between 6 and 8 Å. In Run 1 we observe after 550 ns a short period of about 100 ns where this distance fluctuates between 8 and 12 Å, which is due to partial unbinding as we will show later on. In run 2 the peptide unbinds, i.e., distance ligand–binding site > 10 Å, after 200 ns and in run 3 unbinding occurs after more than 1 μs.

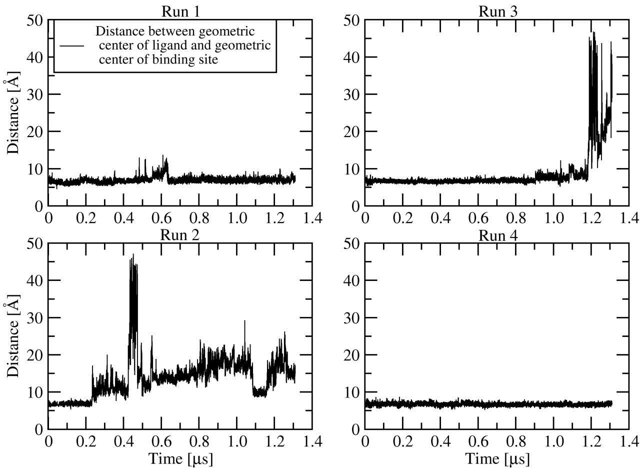


Figure 3.2: **Unbinding.** The four panels show the distance between the geometric center of the C_α atoms of the ligand and the geometric center of the binding site C_α's (defined as C_α's of residues 19-24 of beta-strand $\beta 2$ and 71-79 of helix $\alpha 2$) for the four executed MD runs.

In order to study the role of individual interactions during unbinding we consider only the three runs where at least partial unbinding was observed, i.e., runs 1, 2 and 3. The four native backbone hydrogen bonds emerging from the crystal structure (3LNY) ordered from C- to N-terminus of the ligand are: Val0 N – Ile20 O, Ser-2 O – Val22 N, Ser-2 N – Val22 O and Gln-4 O – Gly24 N (see Figure 3.1). The time series of the distances between donor and acceptor atoms of these hydrogen bonds show that there is indeed a difference in stability (see Figure 3.3). In fact, Gln-4 O – Gly24 N is the most unstable hydrogen bond among the four examined, and it is often not formed when the peptide is bound. This reflects the fact that also in the X-ray structure the donor/acceptor distance of this hydrogen bond is rather large (≈ 3.9 Å). Furthermore, Figure 3.3 shows clearly that in run 1 and 3 Gln-4 O – Gly24 N and Ser-2 N – Val22 O are completely disrupted about 200 ns prior to disruption of Val0 N – Ile20 O and Ser-2 O – Val22 N. Thus, unbinding seems to proceed from the N- to the C-terminus of the peptide. In addition, run 3 shows that during the unbinding process the two C-terminal hydrogen bonds reform after rupture before the peptide unbinds completely. On the other hand, in run 1 after rupture of all the four native backbone hydrogen bonds, the four distance time series consolidate again, which means that the peptide is bound in another conformation or at another site of the protein. We will see later that it is the carboxylate group that provides the crucial interactions. Run 2 shows a slightly different picture. There, unbinding proceeds rather rapidly compared to the other two runs. But also there the donor/acceptor distances between the 2 C-terminal hydrogen bonds grow slower than the ones of the two N-terminal hydrogen bonds. Rupture of the N-terminal hydrogen bonds about 200 ns prior to rupture of the C-terminal ones in run 1 and 3 indicates that there is a metastable binding mode where the ligand is bound just via its C-terminal part, which is in agreement with the studies of Elkins et al. and Long et al. [1, 4]. Furthermore, it supports the data collected in the study of Per Jemth and coworkers [5] where amide-ester mutations have been used at specific residue positions of the ligand in order to probe the relevancy of the specific hydrogen bonds. The experiments on four different PDZ domains showed that in general mutations at the C-terminus of the peptide resulted in a greater loss of binding affinity than mutations in the N-terminal part.

A qualitative view of the unbinding events is given in Figure 3.4. The conformers of the ligand sampled in the MD runs are shown every 20 ns and colored according to time, starting from red and ending with blue. Although this is just a qualitative impression, it confirms the finding that the N-terminal part of the peptide unbinds first and that there is a metastable state where the peptide is bound only with its C-terminus.

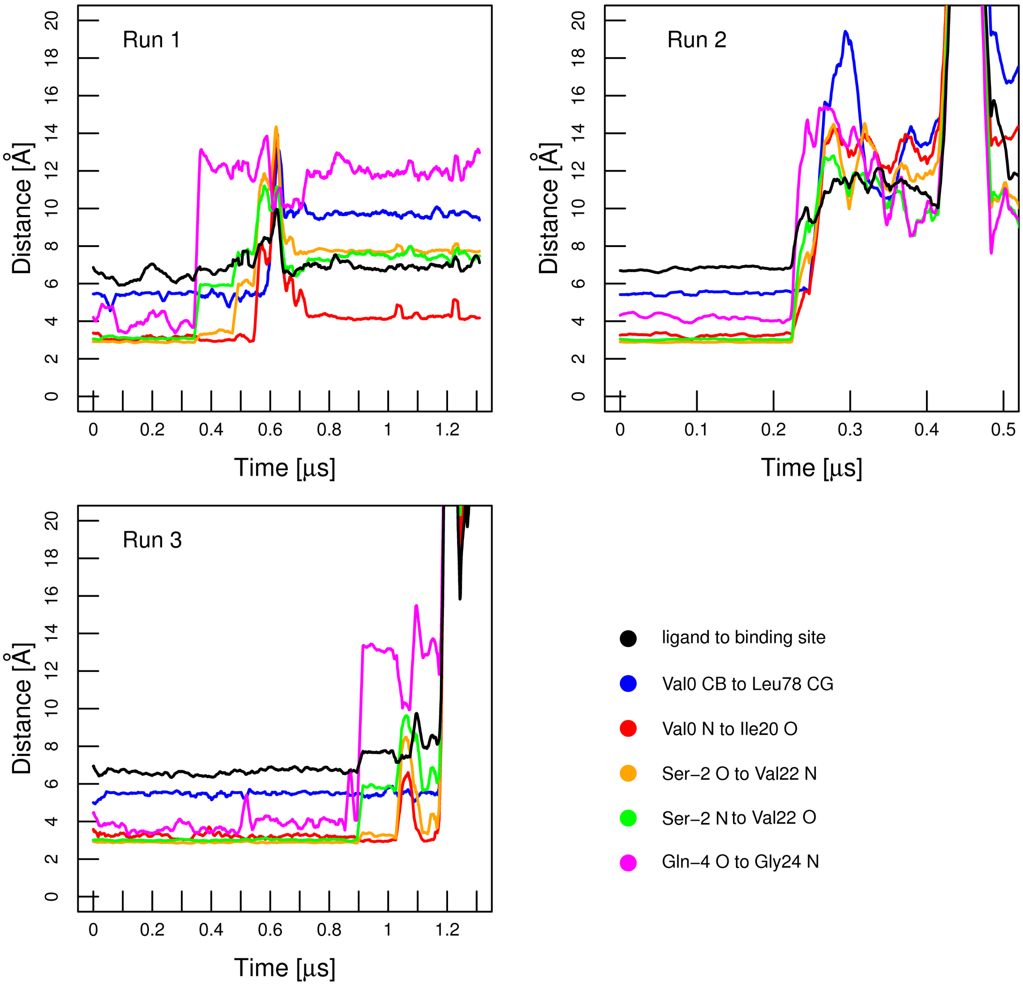


Figure 3.3: **Timeseries of backbone intermolecular hydrogen bonds.** The timeseries of distance between donor/acceptor atoms are shown in red, orange, green and magenta, for the runs with (partial) unbinding events. In order to show the rupture of hydrophobic contact between the side chain of Val0 and the hydrophobic core of the binding site cavity the distance between $C_\beta(CB)$ of Val0 and C_γ (CG) of Leu78 is displayed (blue). A further general indication for unbinding is the distance between ligand and binding site that was computed as in Figure 3.2 (black line). All timeseries are shown as running averages that were computed over time intervals of 20 ns.

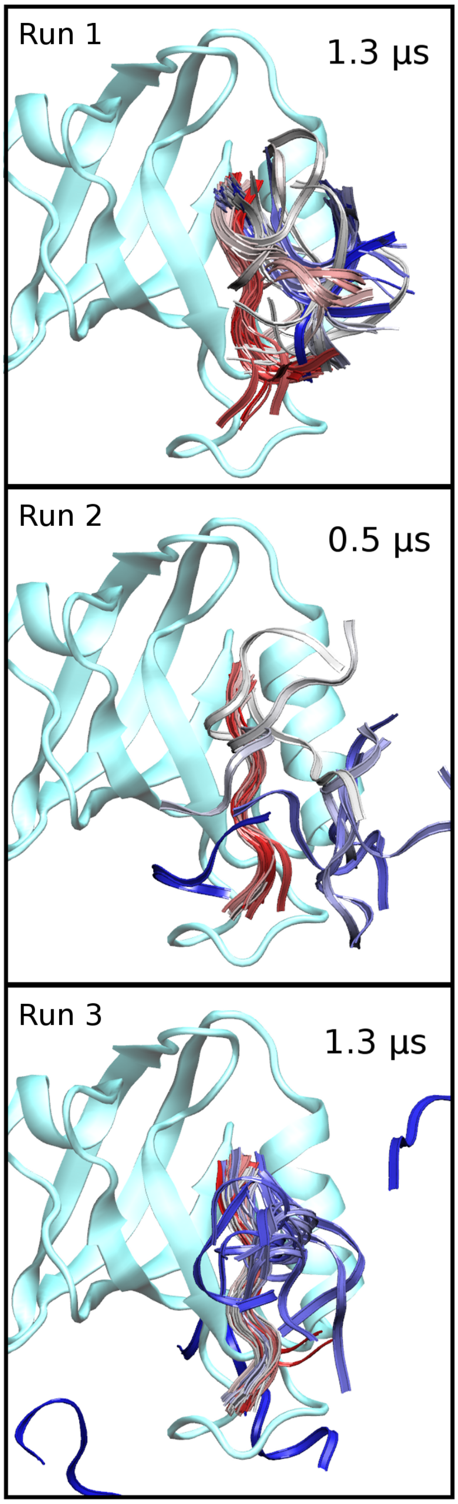


Figure 3.4: **Conformational ensemble of the ligand along the trajectories.** For the three trajectories with a (partial) unbinding event, the conformations adopted by the ligand along time are shown in ribbon representation. The ligand is shown every 20 ns for each of the three runs, the coloring reflects the time, from red to blue, i.e., red corresponds to 0 μs and blue to 1.3, resp. 0.5 μs (run 1 and 3, resp. 2). The trajectory was aligned on the backbone of the first snapshot. Note that some unbound conformations might not be visible due to clipping plane or zoom.

As seen above, in run 1 the peptide ligand readopts a stable confirmation after unbinding. Figure 3.4 indicates that the C-terminus is still close to its native binding position but somehow rotated around its own axis. Figure 3.5 shows distances representing interactions of the carboxylate group of the ligand and the PDZ domain. The native interactions (i.e., the ones found in the X-ray structure) are the ones with Leu18 N, Gly19 N, Ile20 N (see Figure 3.1 red dashed lines) and Ser17 OG. The simulations confirm that the carboxylate group is mostly interacting with this four residues during the time when it is bound. However, we found another important interaction that can be sporadically formed, which is the one between the peptide carboxylate group and the guanidinium group of Arg79, a residue on the helix $\alpha2$. In fact, this is the key interaction that keeps the peptide bound to the protein after rupture of the native interactions in run 1. When this interaction is formed the peptide is rotated around its own axis with respect to the native bound state such that the side chain of Val0 points towards the solvent.

Naturally, due to the small number of simulations and thus observed unbinding events we can only give a qualitative result here. For a quantitative analysis of unbinding many more runs are needed.

3.3 Methods

The explicit solvent MD simulations were carried out with GROMACS [6] using the CHARMM22 all-hydrogen force field [7] and the TIP3P water model [8]. The coordinates of the complex between PDZ2 and the 6-residue C-terminal peptide of RA-GEF-2 were downloaded from the protein data bank (PDB code 3LNY) and used as starting structures. To reproduce neutral pH conditions the side chains of aspartates and glutamates as well as the C-termini of peptide and protein were negatively charged, those of lysines and arginines as well as the N-terminus of the protein were positively charged, and histidines were considered neutral. The N-terminus of the peptide was acetylated. The protein was immersed in a truncated octahedral box of preequilibrated water molecules.

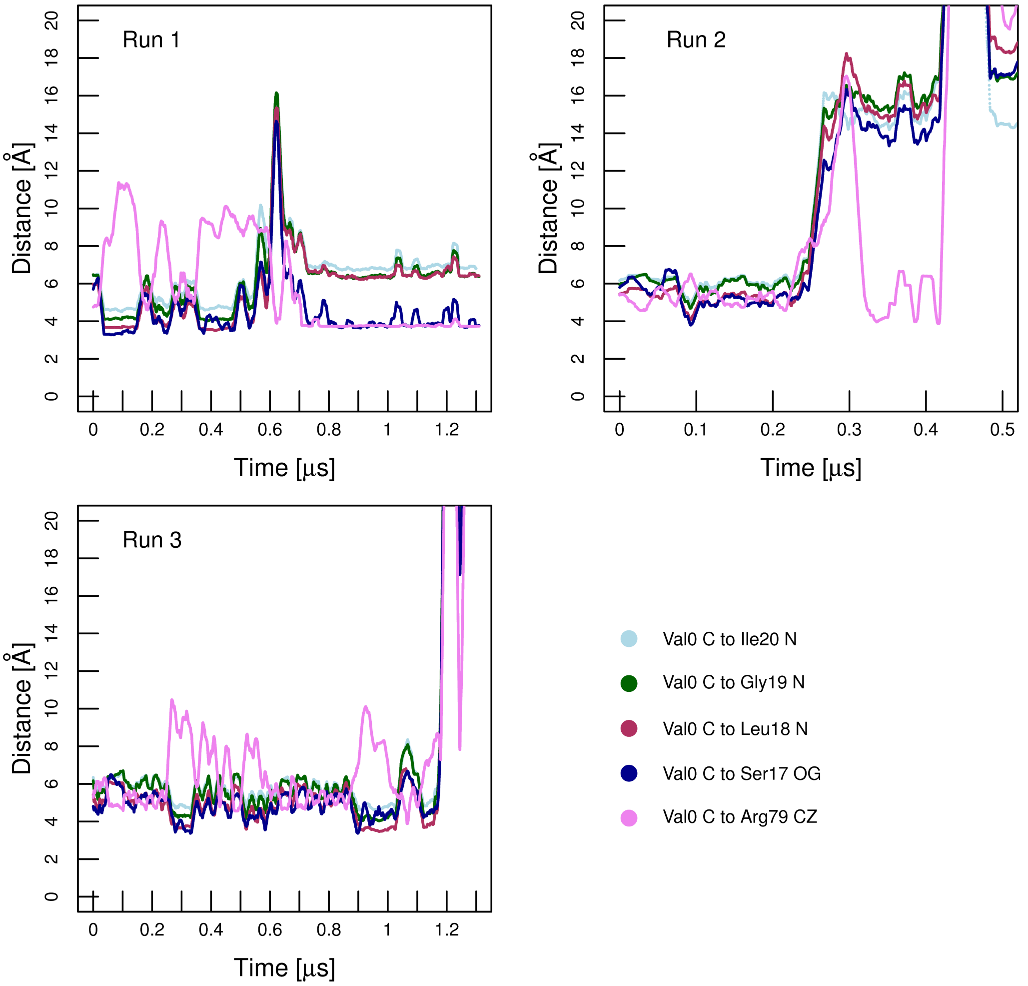


Figure 3.5: **Timeseries of carboxylate intermolecular interactions.** The timeseries of distance between atom C of Val0 (since OT1 and OT2 are symmetric) and amide nitrogen atoms providing anchoring as well as between C of Val0 and Ser17 OG and Arg79 CZ are shown in light blue, dark green, maroon, dark blue and violet, respectively, for the runs with (partial) unbinding events. All timeseries are shown as running averages that were computed over time intervals of 20 ns.

The size of the box was chosen to have a minimal distance of 12 Å between the boundary and any atom of the protein. The simulation box contained two potassium ions to compensate for the net negative charge of the PDZ2/peptide complex.

Periodic boundary conditions were applied, long-range electrostatic interactions were treated with the Particle Mesh Ewald method [9], and the van der Waals interactions were truncated at a cutoff of 10 Å using a switch function starting from 8 Å. The MD simulations were performed at constant temperature (298 K), kept constant by an external bath with velocity rescaling [10], and constant pressure (1 atm) with an integration step of 2 fs and a saving frequency of 2 ps. After equilibration and heating phases the production runs were started using different seeds for the initial distribution of velocities.

The analysis of the MD trajectories was carried out with the MD analysis tool WORDOM [11, 12].

Bibliography

[1] Jonathan M. Elkins, Carina Gileadi, Leela Shrestha, Claire Phillips, Jing Wang, João R.C. Muniz, and Declan A. Doyle. Unusual binding interactions in PDZ domain crystal structures help explain binding mechanisms. *Protein Sci*, 19(4):731–741, 2010.

[2] Ho-Jin Lee and Jie Zheng. PDZ domains and their binding partners: structure, specificity, and modification. *Cell Commun. Signaling*, 8(1):8, 2010.

[3] Z. Songyang, A. S. Fanning, C. Fu, J. Xu, S. M. Marfatia, A. H. Chishti, A. Crompton, A. C. Chan, J. M. Anderson, and L. C. Cantley. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science*, 275(5296):73–77, 1997.

[4] Jia-Fu Long, Wei Feng, Rui Wang, Ling-Nga Chan, Fanny C F Ip, Jun Xia, Nancy Y Ip, and Mingjie Zhang. Autoinhibition of X11/Mint scaffold proteins revealed by the closed conformation of the PDZ tandem. *Nat Struct Mol Biol*, 12(8):722–728, 2005.

[5] Jonas N. N. Eildal, Greta Hultqvist, Thomas Balle, Nicolai Stuhr-Hansen, Shahrokh Padrah, Stefano Gianni, Kristian Strømgaard, and Per Jemth. Probing the role of backbone hydrogen bonds in protein-peptide interactions by amide-to-ester mutations. *Journal of the American Chemical Society*, 0(just accepted):10.1021/ja402875h, 2013.

[6] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E Mark, and Herman J C Berendsen. Gromacs: fast, flexible, and free. *J Comput Chem*, 26(16):1701–1718, 2005.

[7] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Jr. Dunbrack, and J. D. Evanseck et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B*, 102:3586–3616, 1998.

[8] W. L. Jorgensen, J. Chandrasekhar, J. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*, 79:926–935, 1983.

[9] T. Darden, D. York, and L. G. Pedersen. Particle mesh Ewald: an Nlog(N) method for Ewald sums in large systems. *J Chem Phys*, 98:10089, 1993.

[10] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126(1):014101, 2007.

[11] Michele Seeber, Marco Cecchini, Francesco Rao, Giovanni Settanni, and Amedeo Caflisch. Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics*, 23(19):2625–2627, 2007.

[12] Michele Seeber, Angelo Felline, Francesco Raimondi, Stefanie Muff, Ran Friedman, Francesco Rao, Amedeo Caflisch, and Francesca Fanelli. Wordom: a user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. *J Comput Chem*, 32(6):1183–1194, 2011.

Chapter 4

Kinetic response of a photo-perturbed allosteric protein

Buchli, B., Waldauer, S.A., Walser, R., Donten, M., Pfister, R., Blöchliger, N., Steiner, S., Caflisch, A., Zerbe, O. and Hamm, P.; *Proceedings of the National Academy of Sciences*, 110(29): 11725-11730, 2013

Kinetic response of a photoperturbed allosteric protein

Brigitte Buchli[a,1], **Steven A. Waldauer**[a,1], **Reto Walser**[a,1], **Mateusz L. Donten**[a], **Rolf Pfister**[a], **Nicolas Blöchliger**[b], **Sandra Steiner**[b], **Amedeo Caflisch**[b], **Oliver Zerbe**[a], and **Peter Hamm**[a,2]

Departments of [a]Chemistry and [b]Biochemistry, University of Zurich, CH-8057 Zurich, Switzerland

Edited by Hans Frauenfelder, Los Alamos National Laboratory, Los Alamos, NM, and approved June 3, 2013 (received for review April 3, 2013)

By covalently linking an azobenzene photoswitch across the binding groove of a PDZ domain, a conformational transition, similar to the one occurring upon ligand binding to the unmodified domain, can be initiated on a picosecond timescale by a laser pulse. The protein structures have been characterized in the two photoswitch states through NMR spectroscopy and the transition between them through ultrafast IR spectroscopy and molecular dynamics simulations. The binding groove opens on a 100-ns timescale in a highly nonexponential manner, and the molecular dynamics simulations suggest that the process is governed by the rearrangement of the water network on the protein surface. We propose this rearrangement of the water network to be another possible mechanism of allostery.

Subtle conformational transitions within the folded state of highly structured proteins are often an integral aspect in their functional mechanism. These conformational transitions can occur as a result of different events, such as ligand binding, covalent modification (e.g., phosphorylation), or proteolytic cleavage. When an event or perturbation at one site in a protein changes the enzymatic activity or the binding affinity to a ligand at another distant site, this process can be described as allostery. Hemoglobin has long served as the prototypical example to study this effect, where the binding of an oxygen in one subunit modifies the affinity of binding oxygen in another subunit (1, 2). The traditional models of allostery developed by Monod et al. (3) and Koshland et al. (4) attribute allosteric effects to conformational changes by which the allosteric binding site communicates with the distant active site. There is, however, increasing evidence that allostery can be mediated also without any conformational change, relying purely on changes in internal protein dynamics (5).

PDZ domains are an important class of protein interaction modules that have been studied extensively in the context of allostery. They are found in a large variety of proteins and generally bind the C termini of their targets (6–11). As scaffolding domains, they are molecular switches that play a central role in signal transduction. For several PDZ domain proteins, allosteric interactions are an important regulatory mechanism (10, 12–15).

NMR spectroscopy has been particularly useful to elucidate the equilibrium dynamics of proteins on various timescales. The notion of allostery mediated through a change in dynamic properties was corroborated by a study of the third PDZ domain from the PSD-95/SAP90 protein (16). This protein contains an additional C-terminal α-helix (α_3), which shows no direct interaction with the peptide ligand. Removal of α_3 has a negligible effect on the structure of the PDZ core domain; however, it does lead to a large decrease in ligand binding affinity, which was shown to be entirely entropic in nature. Other studies have identified changes in (conformational) entropy of both backbone (17) and side-chain (18) dynamics in other systems to give rise to allosteric effects.

Here, we focus on the second PDZ (PDZ2) domain from human tyrosine-phosphatase 1E (hPTP1E), which has been demonstrated to possess allosteric properties (19). The PDZ domain is a small 96-residue protein with a binding groove between the α_2-helix and the β_2-strand (Fig. 1*B*). As mentioned previously, side-chain dynamics in contiguous sectors spanning the whole protein have been the proposed allosteric mechanism (19–21), but in this case, ligand binding also results in a small but significant structural change, albeit being quite small (22–24) (Fig. 1*B* and Table 1). Furthermore, a number of computational and experimental studies have addressed signal transduction pathways in the PDZ2 model system (25–34).

Allostery is the propagation of a signal between two sites of a protein. Most of the investigations so far have addressed the question of what that signal might be, e.g., a structural change vs. a change in dynamic properties. Even less is known of how such a signal propagates. Whereas NMR spectroscopy is extremely powerful in elucidating equilibrium dynamics on many timescales through relaxation experiments, its inherent time resolution is rather limited in studies of nonequilibrium processes, such as signal propagation. Transient IR spectroscopy, in contrast, provides essentially unlimited (i.e., picosecond) direct time resolution together with still significant chemical selectivity.

To make the best use of the high time resolution, it is crucial to be able to perturb the system locally and with a short laser pulse. Ideally, one would phototrigger the association or dissociation of a ligand. Here, we take an experimentally more feasible approach by covalently linking an azobenzene derivative across the binding groove, which can be switched between *cis* and *trans* isomers with a light of different wavelengths (Fig. 1*A*) (35–38). We carefully designed the system such that the structural perturbation upon isomerization of the photoswitch mimics that upon ligand binding, as is discussed in the next section. Subsequently, we use transient IR spectroscopy to investigate the nonequilbrium transition between both states and finally use nonequilibrium molecular dynamics (MD) simulations to complement the experimental results with atomistic detail.

Structural Characterization

In close analogy to ref. 39, we identified the surface-exposed amino acid pair Ser21 and Glu76 as anchor points for the photoswitch because their C_α-C_α distances in the apo and holo forms closely match those of the photoswitch length in its two configurations. Furthermore, these two residues face across the binding groove at the center. The residues were mutated to cysteines to which the photoswitch was covalently coupled (Fig. S1) (40). When searching for the anchor points, we used the NMR ensemble of structures of the PDZ2 domain [holo, 1D5G, ref. 23); apo, 3PDZ, ref. 22)], which reveals a rather large 1.5-Å change

---

Author contributions: A.C., O.Z., and P.H. designed research; B.B., S.A.W., R.W., M.L.D., N.B., S.S., and P.H. performed research; B.B., S.A.W., R.W., and R.P. contributed new reagents/analytic tools; B.B., S.A.W., R.W., O.Z., and P.H. analyzed data; and B.B., S.A.W., R.W., and P.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The NMR, atomic coordinates, chemical shifts, and restraints have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 2M0Z and 2M10) and the BioMagResBank, www.bmrb.wisc.edu (accession nos. 18833 and 18834).

[1] B.B., S.A.W., and R.W. contributed equally to this work.

[2] To whom correspondence should be addressed. E-mail: phamm@pci.uzh.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1306323110/-/DCSupplemental.

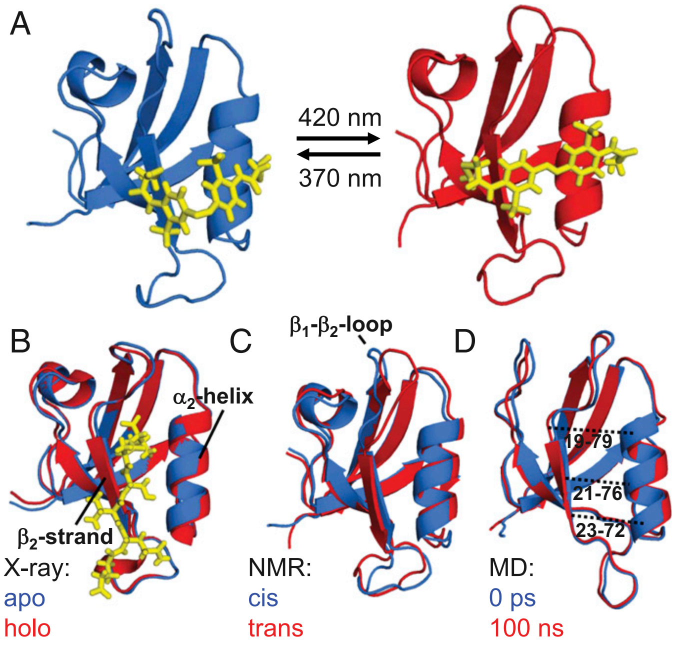


Fig. 1. (*A*) Averaged NMR structures of the photoswitchable PDZ2 domain with the photoswitch (yellow) in the *cis* (*Left*, PDB ID 2M0Z) and *trans* (*Right*, PDB ID 2M10) conformations. (*B*) Overlays of the apo (blue) and holo (red) X-ray structures [3LNX and 3LNY (24)] together with the ligand (the Ras guanine nucleotide exchange factor 2 C-terminal peptide, in yellow) in the latter case. (*C*) The NMR structures with the photoswitch *in cis* (blue) and *in trans* (red) and (*D*) the averaged MD structures with the photoswitch *in cis* (blue) and 100 ns after switching into *trans* (red). For clarity, the photoswitch is not shown in *C* and *D*. The dotted lines in *D* indicate the C_α-C_α distances shown in Fig. 4*B*.

(on average) for the Cα(21)-Cα(76) distance. Recently, X-ray crystallography (3LNX and 3LNY, ref. 24) has shown, however, that this change is about a factor of 2 smaller (i.e., 0.8 Å). The fact that the linker could nevertheless be successfully coupled and resulted in a stable, well-folded protein indicates that the possible structural perturbation is well tolerated.

The two equilibrium structures of the PDZ2 domain with the photoswitch in the *cis* and *trans* configurations were determined by NMR spectroscopy [see *SI Text* for details, Protein Data Bank (PDB) IDs 2M0Z and 2M10]. Whereas the *trans* form of the azobenzene photoswitch is predominant (∼90%) after equilibration in the dark, the *cis* state was generated and maintained (>90%) by continuously illuminating the sample inside the spectrometer with a 370-nm laser coupled to a glass fiber leading into the NMR tube (for further information on the design see *SI Text*), similar to that described in previous work (41, 42). Both forms are stably folded and structurally similar to the corresponding X-ray structures (24), as can be seen in Fig. 1 *B* and *C*. More quantitatively, the rmsd of the NMR structure in *cis* compared with the X-ray structure in the apo form is 1.0 Å and that of *trans* to the holo form is 1.1 Å (considering all backbone heavy atoms of only the regions with defined secondary structure). These values include the uncertainties in the structure determination, the different environments (crystal vs. solvent), and the fact that the molecule is modified by the photoswitch. Also the adaptation of the structure upon isomerization of the photoswitch is comparable to that upon ligand binding. Table 1 lists the rmsds between the corresponding structures for all secondary structure elements and for residues of the α_2-helix and β_2-strand only, with the latter constituting the binding groove. Our construct does slightly overemphasize the conformational perturbation, but the overall agreement is quite reasonable.

Table 1. Structural comparison

rmsd	X-ray (24): apo→holo	NMR: *cis*→*trans*	MD: *cis*→*trans*
All secondary	0.34	0.92	0.46
α_2 and β_2	0.41	0.80	0.62

Structural difference of the apo vs. the holo form deduced from the X-ray structures (3LNX and 3LNY, ref. 24) or the cis vs. the trans conformer from the NMR structures and the MD simulations, respectively. The first row reports the rmsd (in angstroms) when considering all backbone atoms of regions with defined secondary structure and the second row that when considering only the α_2-helix and the β_2-strand.

Transient IR Spectroscopy

Having established the equilibrium structures of the starting and final states by NMR spectroscopy, we now turn to IR spectroscopy. Fig. 2*A* displays stationary FTIR spectra in the region of the amide I band, which is a sensitive probe of the structure of the protein backbone. All IR experiments have been performed in a fully hydrated state [50 mM borate buffer (pH 8.5) and 150 mM NaCl, lyophilized and dissolved in D_2O at 1.3 mM concentration].

A difference signal upon switching *cis* → *trans* or *trans* → *cis* is clearly visible (Fig. 2*A*, red and green) with an intensity of about 1/50th that of the absolute amide I band (Fig. 2, blue), indicating that small changes in the protein backbone do indeed occur. The *trans* → *cis* difference spectrum, induced by illuminating a dark-adapted sample at 370 nm, and the *cis* → *trans* difference spectrum, after switching off the 370-nm light and subsequent relaxation back to *trans* in ∼30 min, are mirror images of each other (Fig. 2*A*, green and red lines), confirming that the molecule can be switched reversibly. In the next step, to observe the

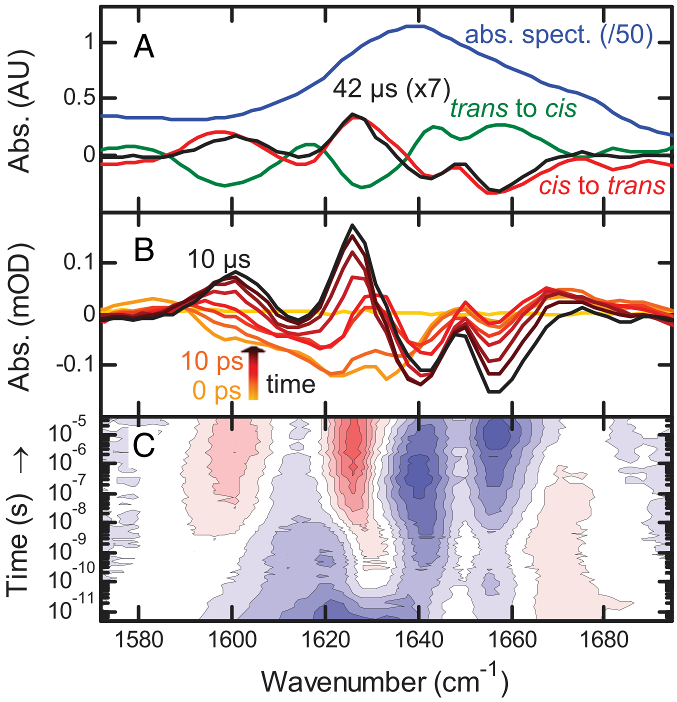


Fig. 2. (*A*) Absolute (photoswitch *in trans*, blue, downscaled by 50) and difference FTIR spectra (red and green) compared with the transient spectrum at 42 µs (black, upscaled by 7). (*B*) Transient difference spectra at −1 ns (yellow), 0 ps (light orange), and from 10 ps to 10 µs by decade (orange to black). (*C*) Contour plot of the IR response. Red indicates induced absorption, blue indicates a bleach, and contour lines are equally spaced.

transition in time-resolved experiments, we prepared all samples in the *cis* configuration by continuously illuminating them at 370 nm with an excess amount of light and then initiated the *cis* → *trans* transition with a picosecond 420-nm pulse (*Materials and Methods*). At long times (42 µs), the transient IR response in the amide I region closely resembles the shape of the steady state (Fig. 2*A*, black vs. red line), indicating that the structural transition is essentially complete after this time. Its amplitude its about one-seventh of the latter, allowing one to estimate the combined excitation probability and isomerization yield.

In reaching the final state, the data reveal a complex evolution over many orders of magnitude in time with bands that both change in intensity and shift in frequency (Fig. 2 *B* and *C*). No physically meaningful model with a limited number of discrete intermediate states could be identified to which we could globally fit the data. Similar observations have been made for downhill folding (43, 44), with different kinetic responses for different spectroscopic observables. This behavior is indicative of a continuous transition between initial and final states without any significant barriers on the pathway.

Three major phases of the overall process can nonetheless be identified. We illustrate these phases in Fig. 3 for the amide I band, choosing 1,640 $\mathrm{cm^{-1}}$ as probe wavelength (red circles), and a strong band at 1,491 $\mathrm{cm^{-1}}$ originating from the photoswitch linked to the PDZ2 domain (Fig. 3, green circles). This band is the amide II vibration of the amide unit of the linker connecting the azobenzene moiety to the protein (Fig. S1*A*). For comparison, the response of the same band of the unlinked photoswitch is shown as well (Fig. 3, blue circles; see Fig. S2, for the complete data). Phase I, clearly visible in the photoswitch bands in both the linked and unlinked form, is initiated by the absorption of a 420-nm photon and the subsequent ultrafast isomerization of the azobenzene moiety. This results in the deposition of a large amount of energy into the vibrational degrees of freedom of the molecule. The vibrational energy appears as heat and results in a broad IR signal, decaying within a few 10s of picoseconds as it quickly dissipates into the solvent (45). The heat signal happens to be zero for the amide I band at 1,640 $\mathrm{cm^{-1}}$, but it is clearly visible at other probe wavelengths (Fig. 2*C*).

Subsequently, in phase II, the band of the photoswitch linked to the PDZ2 domain evolves in a highly nonexponential manner until about 100 ns (Fig. 3, green circles), significantly slower than that of the photoswitch alone (Fig. 3, blue circles).

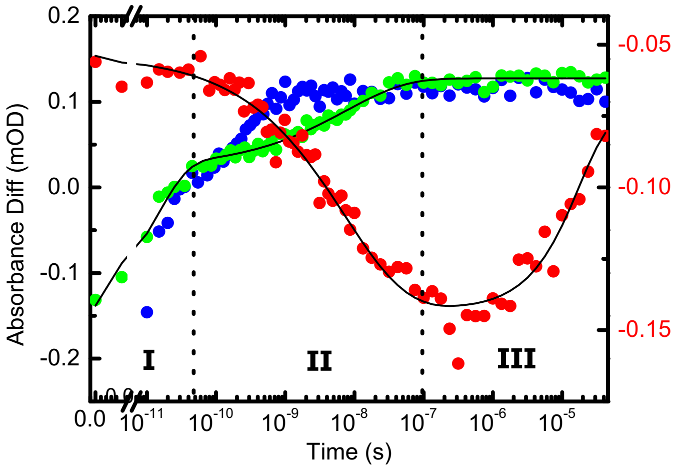


Fig. 3. Amide I response at a selected wavelength of 1,640 $\mathrm{cm^{-1}}$ (red, right scale) compared with that of a band at 1,491 $\mathrm{cm^{-1}}$ from the unlinked photoswitch (blue, left scale) and with that of the same band of the photoswitch linked to the PDZ2 domain (green, left scale). The signal of the photoswitchable PDZ2 domain (red and green) is fitted jointly to Eq. **1** (black).

Photoisomerization of the azobenzene moiety is an ultrafast picosecond process (46). The heat signal of the photoswitch linked to the protein appears on the same timescale as that of the unlinked photoswitch (Fig. S2); hence, in terms of crossing from the electronically excited back to the ground state, and thus in terms of the configuration of the central N = N bond, isomerization is equally fast in both cases. When bound to the protein, however, the photoswitch will find itself in a highly strained state after isomerization, because the protein cannot adapt instantaneously. This strain will affect also the vibrational states of the linker connecting the azobenzene moiety to the protein, i.e., the 1,491-$\mathrm{cm^{-1}}$ band. As the protein relaxes, the strain on the photoswitch is slowly released. Hence, the phase II signal of the photoswitch (Fig. 3, green circles) can be thought of as an indirect measure of the perturbation of the binding groove, as the binding groove is cross-linked by the photoswitch. Similar conclusions have been drawn for the electronic response of a similar azo-photoswitch in a smaller peptide (36). The perturbation of the binding groove is also reflected in the amide I band of the protein (Fig. 3, red circles).

Finally, in phase III, the photoswitch signal remains constant, because the binding groove has fully adapted to the perturbation. The amide I signal nevertheless continues to evolve in time. We assume that this signal is related to a slight rearrangement of the protein structure in a region different from the binding groove, for instance of some of the turn regions that are known to be quite flexible. This interpretation is corroborated by signal broadening of NMR resonances from the β_1-β_2 loop.

The two time traces from the photoswitchable PDZ2 domain at 1,640 $\mathrm{cm^{-1}}$ and at 1,491 $\mathrm{cm^{-1}}$ can be fitted jointly to a function (Fig. 3, black lines)

$$q(t)=a_0+a_1 e^{-(t/\tau_1)}+a_2 e^{-(t/\tau_2)^{\beta}}+a_3 e^{-(t/\tau_3)}, \qquad [1]$$

which is composed of a fast exponential contribution for the heat signal decay in phase I (τ_1 = 15 ps), a stretched exponential contribution for the binding groove perturbation in phase II (τ_2 = 7 ns, β = 0.49), and another exponential contribution (τ_3 = 20 µs) for the final relaxation in the amide I band in phase III. In that fit, the time constants and the stretching factor were forced to be the same for both time traces, but the amplitudes were allowed to differ (Table S1). With a stretching factor of β = 0.49, the nonexponential time dependence of phase II is quite pronounced.

Nonequilibrium Molecular Dynamics

To facilitate the understanding of the transition on an atomistic level, we used nonequilibrium MD simulations in explicit water. Starting from an equilibrated *cis* ensemble (with an rmsd to the corresponding NMR structure of 1.4 Å), we launched very many nonequilibrium trajectories by instantaneously switching the potential energy function of the central N = N bond of the photoswitch from one that is stable in *cis* to one that is stable in *trans* (*Materials and Methods*) (36, 38, 47). As these simulations are limited to a maximum simulation time of 100 ns, we focus on phase II in the following discussion, i.e., the perturbation of the binding groove. The overall fold does not change during the first 100 ns after photoswitching (Fig. 1*D*), but the protein backbone is deformed slightly, as expressed by the rmsd that increases to 0.46 Å after 100 ns (Table 1 and Fig. 4*A*). The rmsd is larger when considering the α_2-helix and the β_2-strand only (0.62 Å), emphasizing that most of the structural changes occur at the binding groove on this timescale. As a function of time, the rmsd jumps relatively rapidly within the first 1 ps. It then continues to grow in a highly nonexponential manner, covering all orders of magnitude in time considered in this simulation, similar to the experimental observation (Fig. 3, green circles), and in fact it is

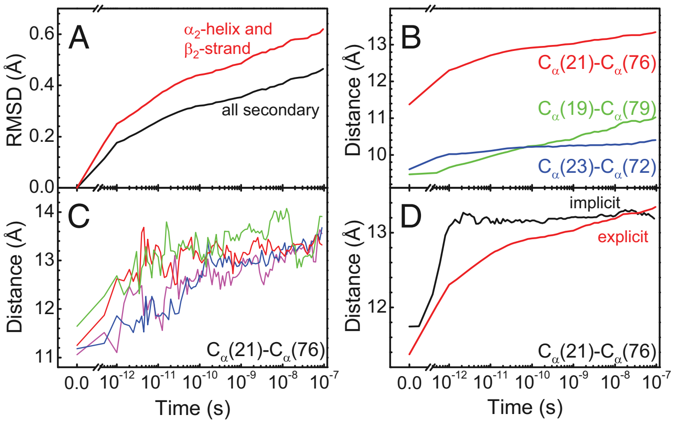


Fig. 4. MD results. (*A*) Time evolution of the rmsd, averaged over an ensemble of nonequilibrium trajectories. The rmsd is relative to the averaged starting structure. (*B*) The same for the C_α-C_α distances across the binding groove, indicated as dotted lines in Fig. 1*D*. (*C*) $C_\alpha(21)$-$C_\alpha(76)$ distance from four typical individual trajectories. *A*–*C* were all deduced from explicit water simulations. (*D*) Comparison of the $C_\alpha(21)$-$C_\alpha(76)$ distance in simulations performed in explicit (red) vs. implicit (black) water.

not finished after 100 ns. We therefore did not attempt to fit the MD data, because a stretched exponential fit becomes robust only if data exist for times long enough so that the signal levels off. No quantitative agreement is expected for the time dependence because, for example, the self-diffusion coefficient of TIP3P (three-site transferrable intermolecular potential) water used in the simulation is more than a factor of 2 higher than the experimental value (48), and water plays an important part in determining the response of the protein (see discussion below). Nevertheless, qualitatively speaking, the response is similar to phase II in Fig. 3 (green circles) [note that phase I is not directly related to any structural process, but rather to a heat signal, which is based on the anharmonicity of the molecule's potential (45) and as such is beyond the MD model]. Furthermore, the amount of structural change obtained from the MD simulation agrees reasonably well with what we find for the NMR structures (keeping in mind that the transition is not quite complete after 100 ns, Table 1).

As a more direct measure of the structural change of the binding groove, we also show in Fig. 4*B* the time evolution of various C_α-C_α distances across the binding groove. The $C_\alpha(21)$-$C_\alpha(76)$ distance, corresponding to the sites that are directly linked by the photoswitch, is perturbed the most and jumps very quickly within the first 1 ps by a significant amount of ∼1 Å. We attribute this initial jump to the direct impact of the isomerization of the photoswitch on the protein backbone. Thereafter the $C_\alpha(21)$-$C_\alpha(76)$ distance again increases steadily in a highly nonexponential manner. The neighboring pairs [Fig. 4*B*, $C_\alpha(19)$-$C_\alpha(79)$ in green and $C_\alpha(23)$-$C_\alpha(72)$ in blue] experience much less of the initial jump in change of distance, because they are not directly affected by the conformational change of the photoswitch, but the nonexponential response at later times is similar.

Nonexponential protein dynamics, modeled either as stretched exponentials or power laws, have been discussed extensively, for instance in the context of ligand (CO) dissociation and rebinding in hemoglobin or myoglobin (49–51) or the fluctuations of the pairwise distance between two sites in a protein (52–54). Two limiting scenarios of nonexponential relaxation kinetics can be distinguished (50): The parallel process is characterized by a distribution of exponential decay processes, originating from a distribution of barrier heights in an inhomogeneous ensemble of proteins. This mechanism applies in the limit when the timescale of the relaxation process is fast compared with the time the protein requires to sample its complete conformational space. In that case, individual single-molecule trajectories would still be a two-state system with either short or long C_α-C_α distances that are separated by one dominant barrier, so that one would observe essentially sudden jumps between these two states with a nonexponential distribution of jump times. Fig. 4*C* shows that we are in the opposite limit. That is, individual single-molecule trajectories essentially follow the averaged one (apart from statistical noise), without big jumps. In accordance with the conclusion already drawn from the experimental results, this observation implies that the transition is continuous without having to surmount any dominant barrier. Such a nonexponential response occurs when protein relaxation is the result of many small events, like defect diffusion that commonly leads to subdiffusive behavior (55).

Fig. 5 indicates what these many small events might be. Shown is a time series of the change of water density at the surface of the protein averaged over all nonequilibrium trajectories. As the photoswitch is isomerizing very quickly in the simulation (<1 ps), the water density changes immediately in the vicinity of the photoswitch (which is located on the right side in the structures in Fig. 5). As time proceeds, this perturbation of the water network travels around the protein and it in fact takes 100 ns until it reaches the back side. This relatively slow propagation of the perturbation of the water network needs to be put in contrast to the residence time of a given single water molecule at the protein surface, which was calculated to be 10–100 ps in the binding groove and 10–30 ps on the outside surface, respectively. Hence, water molecules need to exchange many times before a new equilibrium of the water network around the protein is established.

To gain further insights into the role of water during the structural transition, we repeated the switching simulations with implicit water (Fig. 4*D*, black line), which approximates water as a continuum of the correct dielectric constant, but without the internal degrees of freedom that would provide friction (no additional friction term was added in the Newtonian dynamics used in these simulations). Both states remain stably folded in the implicit water model, but the kinetics of the transition between them are very different from those of the explicit water model. That is, the average $C_\alpha(21)$-$C_\alpha(76)$ distance completes a 1.4-Å jump within ∼1 ps in a ballistic fashion (a minor overshoot is observed) and then stays essentially constant out to 100 ns. Clearly, the implicit water simulation is artificial and not aimed to reveal results that are comparable to reality, but is used to reveal the consequences of the direct impact of water molecules on protein dynamics through the comparison with the explicit water MD simulation. This numerical experiment shows that on the length scale of this conformational transition, the intramolecular potential of the protein alone is not rugged at all and provides no internal friction. In other words, the process is entirely "slaved" by water (51, 56, 57), and water is an integral part of the observed protein dynamics.

Conclusion

In conclusion, we have shown in a closely linked experimental-simulation study that the perturbation of the binding groove of the PDZ2 domain evolves as a continuous subdiffusive process on a 100-ns timescale. The MD simulation (Fig. 4) can essentially quantitatively reproduce the experimentally observed kinetics (Fig. 3, green circles) both in terms of its timescale and in terms of its nonexponential character. This excellent agreement validates the MD simulation, from which atomistic details may then be extracted. We find that the ruggedness of the free energy landscape that governs the dynamics of the binding groove originates entirely from water, whereas the intramolecular potential of the protein is smooth for this small conformational transition. A similar conclusion was drawn for a significantly larger conformational change between a folding intermediate

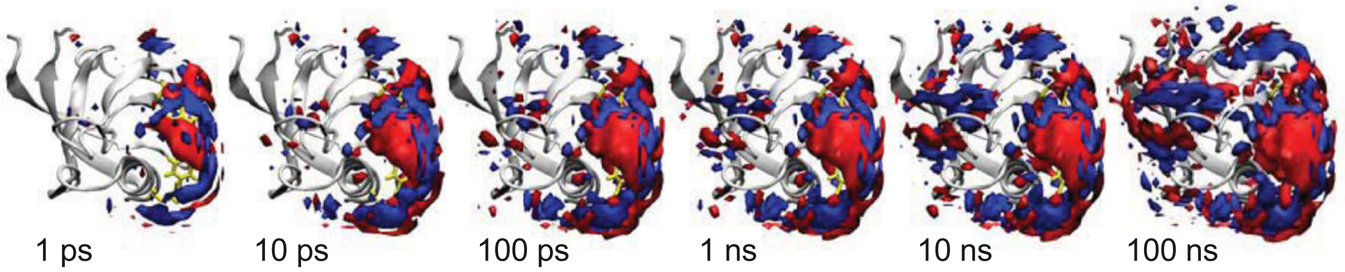


Fig. 5. Change of water density as a function of simulation time, compared with that just before switching. Red depicts increased density and blue decreased density. The contour surfaces correspond to changes of ±0.01 water/Å^3 (for comparison, the bulk water density is ~0.033 water/Å^3). The protein is shown as a gray ribbon and the photoswitch (visible only in part) is shown in yellow. See also Movie S1.

and the native state of a four-helix bundle protein (58). We find that in our system the overall protein response is dominated by the rearrangement of the water network on the protein surface. Interestingly, the perturbation of the water network propagates around the protein within the 100 ns (Fig. 5).

We propose this to be another possible mechanism of allostery, which addresses the question of how the ligand binding site communicates with remote parts of the protein. Although the photoswitch is a rather crude mimic of ligand binding, the peptide ligand will still introduce new partial charges in the binding pocket that will rearrange the water network in its vicinity and eventually also at larger distances. This mechanism would work without any significant structural change of the protein and might even unify the seemingly competing points of view, which explain allostery either as a structural or as a dynamical effect. That is, to the extent that the dynamics of a protein are slaved by water, a change in water structure can affect the dynamics of the protein. Independent from that, phase III in Fig. 3 also hints toward a conformational change of the protein backbone in a region different from the binding groove that also could be responsible for allosteric signaling in a more traditional sense (3, 4).

Materials and Methods

Protein Preparation. The PDZ2 domain (S21C E76C) was expressed from *Escherichia coli*, using standard methods. The photoswitch 3,3′-bis(sulfonato)-4,4′- bis(chloroacetamido)azobenzene (BSBCA) was covalently linked to the two cysteines (40); see *SI Text* and Fig. S3 for details. We learned from mass spectrometry that the protein reacts photochemically with oxygen as well as with the initially used Tris buffer under the influence of the 420-nm laser pulses in the pump-probe experiment, where both presumably bind to the thioether groups of the cysteines linked to the photoswitch (Fig. S4). The experiments were therefore performed in 50 mM borate buffer (pH 8.5) and 150 mM NaCl, lyophilized and dissolved in D_2O, and care was taken to maintain the sample oxygen-free.

Time-Resolved IR Spectroscopy. Two synchronized 1-kHz Ti:sapphire oscillator/regenerative amplifier femtosecond laser systems (Spectra Physics) were used for pump-probe measurements (59). The jitter between both lasers (which effectively determines the time resolution of the experiment) was ~10 ps and the delay could be adjusted up to 42 μs. The frequency-doubled pulses (420 nm, 3 μJ per pulse focused onto an ~200-μm beam diameter in the sample and stretched to ~1 ps to reduce sample deposition on the cuvette windows) of one laser system were used to excite the photoswitch. The IR probe pulses were obtained by sending the output of the second laser system through an optical parametric amplifier (100 fs, center wavenumber 1,635 cm^{-1} or 1,443 cm^{-1}). The sample was circulated in a closed-cycle flow-cell system consisting of a reservoir followed by a CaF_2 sample cell with 50 μm optical path length. The reservoir was continuously illuminated with a 150-mW, 370-nm continuous wave diode laser (CrystaLaser) so that all protein flowing into the sample cell was in the *cis* configuration.

NMR. NMR spectra of PDZ2 with the photoswitch in *trans* were recorded in the dark after equilibration. For all measurements with the photoswitch in the *cis* configuration, the sample was continuously illuminated with the 370-nm cw laser specified above (Fig. S5) coupled into the NMR spectrometer through a custom-fabricated fiber terminated with an extended cylindrical diffuser (Molex). Spectrum assignment was achieved with a standard set of triple-resonance experiments. Structure calculation was performed from NOE distance restraints from ^{15}N- and ^{13}C-resolved NOESY spectra with 75-ms mixing times. NOE data were complemented by amide proton residual dipolar couplings (NH-RDCs) measured in Pf1 bacteriophage and *n*-dodecyl-penta(ethylene glycol)/*n*-hexanol liquid crystalline medium. The 20 conformers with lowest energy of both structures showed no NOE violations bigger than 0.5 Å and showed good Ramachandran plot statistics with only 0.3% of the residues in disallowed regions (for more detailed description of experiments, parameters, and structural statistics see *SI Text* and Tables S2 and S3).

Computational Methods. MD simulations were performed with the Gromacs program package (60) and the Gromacs implementation of the Charmm27 force field (61, 62). The photoswitch was parameterized as in ref. 36. Details of the simulation protocol are given in *SI Text*.

ACKNOWLEDGMENTS. We thank Ben Schuler and his group, in particular Hagen Hofmann, for tremendous help with the protein chemistry; and Andrew Woolley and Gerhard Stock for important discussions; and the Functional Genomics Center Zurich, especially Serge Chesnov, for help with mass spectrometry. This work has primarily been supported by an European Research Council (ERC) Advanced Investigator Grant (DYNALLO) and in part by the Swiss National Science Foundation through the National Center of Competence and Research (NCCR) MUST.

1. Henry ER, Jones CM, Hofrichter J, Eaton WA (1997) Can a two-state MWC allosteric model explain hemoglobin kinetics? *Biochemistry* 36(21):6511–6528.
2. Eaton WA, Henry ER, Hofrichter J, Mozzarelli A (1999) Is cooperative oxygen binding by hemoglobin really understood? *Nat Struct Biol* 6(4):351–358.
3. Monod J, Wyman J, Changeux JP (1965) On the nature of allosteric transitions - a plausible model. *J Mol Biol* 12:88–118.
4. Koshland DE, Jr., Némethy G, Filmer D (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* 5(1): 365–385.
5. Cooper A, Dryden DTF (1984) Allostery without conformational change. A plausible model. *Eur Biophys J* 11(2):103–109.
6. Songyang Z, et al. (1997) Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science* 275(5296):73–77.
7. Daniels DL, Cohen AR, Anderson JM, Brünger AT (1998) Crystal structure of the hCASK PDZ domain reveals the structural basis of class II PDZ domain target recognition. *Nat Struct Biol* 5(4):317–325.
8. Harris BZ, Lim WA (2001) Mechanism and role of PDZ domains in signaling complex assembly. *J Cell Sci* 114(Pt 18):3219–3231.
9. Jemth P, Gianni S (2007) PDZ domains: Folding and binding. *Biochemistry* 46(30): 8701–8708.
10. Lee H-J, Zheng JJ (2010) PDZ domains and their binding partners: Structure, specificity, and modification. *Cell Commun Signal* 8:8.
11. van Ham M, Hendriks W (2003) PDZ domains-glue and guide. *Mol Biol Rep* 30(2):69–82.
12. van den Berk LCJ, et al. (2007) An allosteric intramolecular PDZ-PDZ interaction modulates PTP-BL PDZ2 binding specificity. *Biochemistry* 46(47):13629–13637.
13. Li J, Callaway DJE, Bu Z (2009) Ezrin induces long-range interdomain allostery in the scaffolding protein NHERF1. *J Mol Biol* 392(1):166–180.
14. Wilken C, Kitzing K, Kurzbauer R, Ehrmann M, Clausen T (2004) Crystal structure of the DegS stress sensor: How a PDZ domain recognizes misfolded protein and activates a protease. *Cell* 117(4):483–494.
15. Whitney DS, Peterson FC, Volkman BF (2011) A conformational switch in the CRIB-PDZ module of Par-6. *Structure* 19(11):1711–1722.

16. Petit CM, Zhang J, Sapienza PJ, Fuentes EJ, Lee AL (2009) Hidden dynamic allostery in a PDZ domain. *Proc Natl Acad Sci USA* 106(43):18249–18254.
17. Popovych N, Sun S, Ebright RH, Kalodimos CG (2006) Dynamically driven protein allostery. *Nat Struct Mol Biol* 13(9):831–838.
18. Frederick KK, Marlow MS, Valentine KG, Wand AJ (2007) Conformational entropy in molecular recognition by proteins. *Nature* 448(7151):325–329.
19. Fuentes EJ, Der CJ, Lee AL (2004) Ligand-dependent dynamics and intramolecular signaling in a PDZ domain. *J Mol Biol* 335(4):1105–1115.
20. Fuentes EJ, Gilmore SA, Mauldin RV, Lee AL (2006) Evaluation of energetic and dynamic coupling networks in a PDZ domain protein. *J Mol Biol* 364(3):337–351.
21. Gianni S, et al. (2006) Demonstration of long-range interactions in a PDZ domain by NMR, kinetics, and protein engineering. *Structure* 14(12):1801–1809.
22. Kozlov G, Gehring K, Ekiel I (2000) Solution structure of the PDZ2 domain from human phosphatase hPTP1E and its interactions with C-terminal peptides from the Fas receptor. *Biochemistry* 39(10):2572–2580.
23. Kozlov G, Banville D, Gehring K, Ekiel I (2002) Solution structure of the PDZ2 domain from cytosolic human phosphatase hPTP1E complexed with a peptide reveals contribution of the beta2-beta3 loop to PDZ domain-ligand interactions. *J Mol Biol* 320(4):813–820.
24. Zhang J, et al. (2010) Crystallographic and nuclear magnetic resonance evaluation of the impact of peptide binding to the second PDZ domain of protein tyrosine phosphatase 1*E*. *Biochemistry* 49(43):9280–9291.
25. Ota N, Agard DA (2005) Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion. *J Mol Biol* 351(2):345–354.
26. De Los Rios P, et al. (2005) Functional dynamics of PDZ binding domains: A normal-mode analysis. *Biophys J* 89(1):14–21.
27. Sharp K, Skinner JJ (2006) Pump-probe molecular dynamics as a tool for studying protein motion and long range coupling. *Proteins* 65(2):347–361.
28. Dhulesia A, Gsponer J, Vendruscolo M (2008) Mapping of two networks of residues that exhibit structural and dynamical changes upon binding in a PDZ domain protein. *J Am Chem Soc* 130(28):8931–8939.
29. Kong Y, Karplus M (2009) Signaling pathways of PDZ2 domain: A molecular dynamics interaction correlation analysis. *Proteins* 74(1):145–154.
30. Gerek ZN, Ozkan SB (2011) Change in allosteric network affects binding affinities of PDZ domains: Analysis through perturbation response scanning. *PLoS Comput Biol* 7(10):e1002154.
31. Cilia E, Vuister GW, Lenaerts T (2012) Accurate prediction of the dynamical changes within the second PDZ domain of PTP1e. *PLoS Comput Biol* 8(11):e1002794.
32. Lockless SW, Ranganathan R (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* 286(5438):295–299.
33. Süel GM, Lockless SW, Wall MA, Ranganathan R (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Biol* 10(1):59–69.
34. Chi CN, et al. (2008) Reassessing a sparse energetic network within a single protein domain. *Proc Natl Acad Sci USA* 105(12):4679–4684.
35. Kumita JR, Smart OS, Woolley GA (2000) Photo-control of helix content in a short peptide. *Proc Natl Acad Sci USA* 97(8):3803–3808.
36. Spörlein S, et al. (2002) Ultrafast spectroscopy reveals subnanosecond peptide conformational dynamics and validates molecular dynamics simulation. *Proc Natl Acad Sci USA* 99(12):7998–8002.
37. Rehm S, Lenz MO, Mensch S, Schwalbe H, Wachtveitl J (2005) Ultrafast spectroscopy of a photoswitchable 30-amino acid de novo synthesized peptide. *Chem Phys* 323(1):28–35.
38. Ihalainen JA, et al. (2008) α-Helix folding in the presence of structural constraints. *Proc Natl Acad Sci USA* 105(28):9588–9593.
39. Zhang F, et al. (2009) Structure-based approach to the photocontrol of protein folding. *J Am Chem Soc* 131(6):2283–2289.
40. Burns DC, Zhang F, Woolley GA (2007) Synthesis of 3,3′-bis(sulfonato)-4,4′-bis(chloroacetamido)azobenzene and cysteine cross-linking for photo-control of protein conformation and activity. *Nat Protoc* 2(2):251–258.
41. Rubinstenn G, et al. (1998) Structural and dynamic changes of photoactive yellow protein during its photocycle in solution. *Nat Struct Biol* 5(7):568–570.
42. Kühn T, Schwalbe H (2000) Monitoring the kinetics of ion-dependent protein folding by time-resolved nmr spectroscopy at atomic resolution. *J Am Chem Soc* 122:6169–6174.
43. Garcia-Mira MM, Sadqi M, Fischer N, Sanchez-Ruiz JM, Muñoz V (2002) Experimental identification of downhill protein folding. *Science* 298(5601):2191–2195.
44. Ma H, Gruebele M (2005) Kinetics are probe-dependent during downhill folding of an engineered lambda6-85 protein. *Proc Natl Acad Sci USA* 102(7):2283–2287.
45. Hamm P, Ohline SM, Zinth W (1997) Vibrational cooling after ultrafast photoisomerization of azobenzene measured by femtosecond infrared spectroscopy. *J Chem Phys* 106:519–529.
46. Nägele T, Hoche R, Zinth W, Wachtveitl J (1997) Femtosecond photoisomerization of cis-azobenzene. *Chem Phys Lett* 272:489–495.
47. Nguyen PH, Stock G (2006) Nonequilibrium molecular dynamics simulation of a photoswitchable peptide. *Chem Phys* 323:36–44.
48. Mahoney MW, Jorgensen WL (2001) Diffusion constant of the tip5p model of liquid water. *J Chem Phys* 114:363–366.
49. Lim M, Jackson TA, Anfinrud PA (1993) Nonexponential protein relaxation: Dynamics of conformational change in myoglobin. *Proc Natl Acad Sci USA* 90(12):5801–5804.
50. Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. *Science* 254(5038):1598–1603.
51. Frauenfelder H, et al. (2009) A unified model of protein dynamics. *Proc Natl Acad Sci USA* 106(13):5129–5134.
52. Volk M, et al. (1997) Peptide conformational dynamics and vibrational Stark effects following photoinitiated disulfide cleavage. *J Phys Chem B* 101:8607–8616.
53. Yang H, et al. (2003) Protein conformational dynamics probed by single-molecule electron transfer. *Science* 302(5643):262–266.
54. Milanesi L, et al. (2012) Measurement of energy landscape roughness of folded and unfolded proteins. *Proc Natl Acad Sci USA* 109(48):19563–19568.
55. Jäckle J (1986) Models of the glass transition. *Rep Prog Phys* 49:171–231.
56. Fenimore PW, Frauenfelder H, McMahon BH, Parak FG (2002) Slaving: Solvent fluctuations dominate protein dynamics and functions. *Proc Natl Acad Sci USA* 99(25):16047–16051.
57. Vitkup D, Ringe D, Petsko GA, Karplus M (2000) Solvent mobility and the protein 'glass' transition. *Nat Struct Biol* 7(1):34–38.
58. Sekhar A, Vallurupalli P, Kay LE (2012) Folding of the four-helix bundle FF domain from a compact on-pathway intermediate state is governed predominantly by water motion. *Proc Natl Acad Sci USA* 109(47):19268–19273.
59. Bredenbeck J, Helbing J, Hamm P (2004) Continuous scanning from picoseconds to microseconds in time resolved linear and nonlinear spectroscopy. *Rev Sci Instrum* 75(11):4462.
60. Van Der Spoel D, et al. (2005) GROMACS: Fast, flexible, and free. *J Comput Chem* 26(16):1701–1718.
61. Mackerell AD, Jr., Feig M, Brooks CL, 3rd (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* 25(11):1400–1415.
62. Mackerell AD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102:3586–3616.

Supporting Information

Buchli et al. 10.1073/pnas.1306323110

SI Text

IR-Data

The amide I band was used to observe changes in the protein backbone structure. Additionally, two bands originating from the photoswitch (see Fig. S1*A*), present in the absolute FTIR spectrum in Fig. S2*A* (blue), are sensitive to the photoswitch conformational state. The band at around 1,390 cm^{-1} is a ring mode of the azobenzene and at around 1,490 cm^{-1} is the amide II band of the two photoswitch amide groups (Fig. S1*A*). In the absolute spectra of the photoswitch bound to the protein (Fig. S2*B*, blue), these two modes are hidden behind the stronger amide II band of the protein backbone, but they still are clearly visible in the difference spectra (Fig. S2*B*, red and green), where they appear in a similar way to that for the photoswitch alone (Fig. S2*A*, red and green).

Fig. 3 (blue) of the main text shows a time trace of the photoswitch amide II band. Shortly following the heat signal decay within the first few picoseconds, the spectrum quickly relaxes to closely resemble the steady-state spectrum (Fig. 3 and Fig. S2*A*). When the photoswitch is bound to the second PDZ (PDZ2), both bands overlap with the much broader amide II band of the protein backbone (centered around 1,450 cm^{-1}; Fig. S 1*B*, blue); however, the bands originating from the photoswitch are clearly discernible in the difference spectrum. The kinetics are significantly decelerated by the counterstrain the protein imposes onto the photoswitch.

Table S1 summarizes all parameters obtained from the joint fit of the data in Fig. 3.

SI Materials and Methods

Cloning. The PDZ2 gene, containing the two mutations S21C and E76C, was synthesized and cloned into a pET30a(+)vector (EZbiolab). A human rhinovirus 3C (HRV 3C) cleavage site was added between the N-terminal hexahistidine tag and the protein gene by site-directed mutagenesis (QuikChange; Stratagene). To facilitate detection and quantification of the protein vs. the photoswitchable linker, a Trp was inserted between the His tag and the protease cleavage site (Fig. S2*B*).

Protein Expression and Purification. For IR spectroscopy measurements, the PDZ2 domain was expressed in *Escherichia coli* BL21(DE3) in Luria–Bertani medium. For NMR measurements, M9 minimal medium was supplemented with ^{13}C-glucose and/or ^{15}N-NH_4Cl. Fractional isotope labeling as described by Neri et al. (2) was used for obtaining stereospecific assignments for the Val and Leu methyl groups. The protein was purified from inclusion bodies with a HisPrep column (GE Healthcare Life Sciences) in 20 mM Tris·HCl, 6 M GdmHCl, and 10 mM imidazole, pH 8, and eluted with an imidazole gradient. Residual nickel was removed from the protein by incubation with 40 mM EDTA at 4 °C for at least 12 h.

Linking the Photoswitch to PDZ2. The 3,3′-bis(sulfonato)-4,4′-bis(chloroacetamido)azobenzene (BSBCA) (Fig. S1*A*) was synthesized according to the protocol from Burns et al. (1). The process to covalently link the photoswitch to two surface-exposed Cys residues, as outlined in the protocol, was modified to the following: Disulfide bridges were reduced by adding 50 mM DTT to the protein solution and incubating it for at least 1 h at room temperature. The protein was then refolded by desalting in 50 mM Tris·HCl, pH 8.5, with a HiPrep column (GE Healthcare Life Sciences), simultaneously removing the reducing agent. To prevent the reformation of disulfide bonds, fraction collection was performed in an oxygen-free (argon) atmosphere. The protein was then diluted with a well-degassed buffer (50 mM Tris·HCl, pH 8.5) to 10 μM before 50 μM of BSBCA was added, still under Ar atmosphere. The reaction vessel was sealed and stirred in the dark for 6 h. The reaction mixture was concentrated using a Vivacell pressure concentrator with a 5-kDa MWCO filter (Sartorius). To remove any surplus BSBCA the protein solution was diluted and reconcentrated twice. The monomer with a single photoswitch correctly linked to both Cys was purified using a MonoQ anion exchange column (GE Healthcare Life Sciences) in 50 mM Tris·HCl, pH 8.5, and eluted with an NaCl gradient.

His Tag Cleavage. After linking the photoswitch (BSBCA) to PDZ2, the His tag was removed with HRV 3C protease. The protease used also contained a His tag for separation from the sample. One milligram protease per 50 mg PDZ2 in 50 mM Tris·HCl, pH 8.5, was incubated at 4 °C for 16 h. The His tag-free PDZ2 was purified with a HisTrap column (GE Healthcare Life Sciences) and the buffer was exchanged by a HiPrep desalting column (GE Healthcare Life Sciences) to the corresponding buffer used for IR or NMR measurements.

Mass Spectrometry. Electrospray ionization (ESI) mass spectra were measured at the Functional Genomics Center Zurich in a mass range between 500 and 3,000 Da. The *m*/*z* data were deconvolved using the MasEnt1 software.

The purity of all samples was verified by SDS/PAGE (Fig. S3) and ESI mass spectrometry (Fig. S4*A*). Additionally, ESI mass spectra were repeated after time-resolved IR measurements. Thereby, we realized that the protein was modified during a measurement in Tris buffer (Fig. S4*B*). However, these modifications could be largely avoided by measuring in borate buffer and under exclusion of oxygen (Fig. S4*C*). All transient IR spectra shown in this work were therefore measured under these conditions.

Fig. S4*B* shows one modification with an increased mass of around 16 Da that can occur multiple times. We attribute this to an oxidation of the protein, most likely at the thioether groups of the cysteines linked to the photoswitch. A small part of the protein was already oxidized before the measurement (Fig. S4*A*). However, during a measurement the oxidation significantly increased. A second modification with an increased mass of 135 Da we attribute to a reaction of the oxidized protein with the Tris buffer. Both modifications increased linearly with the time of a measurement and must have been induced by the high intensity of the 420-nm laser pulses. No decomposition products with a lighter mass were detected.

IR Spectroscopy. For FTIR measurements the protein was desalted into 50 mM Tris·HCl, 150 mM NaCl, pH 8.5. We learned that in this buffer the protein is modified under the influence of 420-nm laser pluses (mass spectra, Fig. S4). Therefore, we used 50 mM borate buffer and 150 mM NaCl, pH 8.5, for time-resolved measurements. The protein was concentrated in an Amicon Ultra 3-kDa MWCO centrifugal filter device (Millipore Corporation) to 1.3 mM, lyophilized, and dissolved in D_2O. Incubation in D_2O for 2 h at room temperature before the measurements eliminated H/D exchange during experiments. FTIR spectra were measured on a Bruker Tensor 27 FTIR spectrometer in a 50-μm or 100-μm path-length sample cell with CaF_2 windows, either in the dark or under illumination with a 150-mW, 370-nm continuous wave (cw) diode laser (CrystaLaser). The unlinked photoswitch (Figs. 1*A* and 3) was measured in the same buffer and at 1.1 mM concentration.

All FTIR and transient IR spectra were measured at room temperature (21–22 °C).

NMR Spectroscopy. For NMR spectroscopy the samples (in 50 mM sodium phosphate, 150 mM NaCl, pH 6.8) were concentrated to 0.75 mM. The buffer was exchanged (50 mM sodium phosphate, pH 6.8, 150 mM NaCl, 10% D_2O) in two successive rounds of 1:1 dilution and concentration. For measurements with the photoswitch *in cis* a 150-mW, 370-nm cw diode laser (CrystaLaser) coupled to a fiber with a custom-designed inline radial illumination probe (Polymicro Technologies) (Fig. S5) was used.

Proton chemical shifts were calibrated to the water signal and nitrogen shifts were referenced indirectly to liquid NH_3 (3). All 2D experiments used TPPI-States for quadrature detection in indirect proton dimensions and gradient-selected coherence selection (echo–antiecho) (4) in combination with sensitivity enhancement schemes in experiments including detection of amide protons.

Resonance Assignment. All NMR experiments were recorded at 25 °C on a Bruker Avance 600- or 700-MHz spectrometer equipped with a cryoprobe. Backbone resonances were identified with the help of CBCA(CO)NH (5), HNCACB (6), and HNCO (7) experiments. Side-chain assignments were obtained using HC(C)H-TOCSY (8, 9) and CC(CO)NH (10) experiments. Aromatic spin systems were linked to the backbone via a CB(CGCD)HD experiment (11). Val and Leu methyl groups were stereospecifically assigned (2). Spectra were processed under the Bruker Topspin 2.1 software and then transferred to Cara (12) for further analysis. The overall backbone/sidechain assignment completeness is 98/85% and 96/84% for the *cis* and *trans* forms, respectively. The lower percentage of backbone resonance assignments of the *trans* form is due to a bigger number of exchange-broadened peaks in proximity to the linker attachment points.

Structure Calculation. ^{15}N- and ^{13}C-edited NOESY (13) spectra with mixing times of 75 ms were used for obtaining NOE structural restraints. Two sets of amide proton residual dipolar couplings (NH-RDCs) (14) were measured for both the *cis* and the *trans* form in 8.5 mg/mL filamentous phage Pf1 (15) (ASLA Biotech) and in liquid crystalline media formed with *n*-dodecyl-penta(ethylene glycol) (C12E5) and *n*-hexanol (16). NOESY peaks were picked with ATNOS/CANDID (17, 18). Torsion angle restraints were generated with TALOS+ (19) and final structures were calculated with CYANA (20, 21). For the *trans* form a number of artificial restraints were imposed on the photoswitch to enforce a *trans*-diazo bond and coplanar aromatic rings whereas for the *cis* form only the *cis*-diazo bond was enforced (Table S2).

Structure Refinement. Structures were refined in TIP3P (three-site transferrable intermolecular potential) explicit water, using the Charmm27 force field as described previously (22–24). Parameters for the BSBCA cross-linker were derived by analogy as for azobenzene (25, 26). The structures of the *cis* and the *trans* form both show the β6/α2 architecture characteristic of PDZ2 domains. The *trans* form suffers from a number of exchange-broadened peaks, in particular in the β1-β2 loop resulting in low numbers of available distance restraints for that region concomitant with a less well-defined structure in that stretch compared with the overall structure. We point out here that the determination of the NMR structures relies on a force field similar to the one used for carrying out the molecular dynamics (MD) simulations and that distance values determined from the two methods might thus be not fully independent.

Structure Validation. Structures were validated with WHATCHECK (27, 28) and PROCHECK-NMR (29) (Table S3).

Data Bank Accession Codes. Coordinates and chemical shifts of the *cis* and *trans* configurations have been deposited in the Protein Data Bank (PDB) and the BioMagResBank (BMRB) data bank under PDB ID codes 2M0Z and 2M10 and BMRB accession nos. 18833 and 18834, respectively.

Computation. MD simulations were performed with the Gromacs program package (30) and the Gromacs implementation of the Charmm27 force field (23, 24), with a timestep of 2 fs, saving time 500 fs, all bonds constrained, isothermal–isobaric (NPT) ensemble at 300 K and 1 bar, with time constants of 0.2 ps and 0.5 ps, respectively, for the thermostat and the barostat. Lennard–Jones interactions were treated with a cutoff of 1.0 nm (switched to zero at 0.9 nm), and the long-range electrostatic forces were approximated by the Particle–Mesh–Ewald approximation. The photoswitch was parameterized as in refs. 25 and 26. To force the photoswitch to be either in the *cis* or in the *trans* configuration, the double-minimum potential for the central C-N = N-C dihedral angle was replaced by a single-minimum potential, and the force constant was increased by a factor of 3 such that the potential around the minimum agrees reasonably well with that of the original double-minimum potential from refs. 31 and 32.

PDB entry 3PDZ was used as a starting structure (33) and mutated to Ser21Cys and Glu76Cys to provide an anchor point for the photoswitch. In addition, a Tyr36Trp mutation was included in accordance with an earlier version of the experimental sequence. We inserted a photoswitch without the sulfonate groups. We were able to synthesize proteins with that version of the photoswitch in small yields and verified that despite its hydrophobicity, the protein was still folded (however, the yields of the synthesis were too small to retrieve enough material for NMR and time-resolved IR spectroscopy).

As in the experiment, we first prepared the photoswitch in its *cis* configuration. The protein was solvated in a box of 5,355 TIP3P water molecules and one Cl^- counter-ion to neutralize the simulation box, minimized with the backbone atoms constrained, and then equilibrated for 1.1 μs. From a subsequent 3-μs *cis*-equilibrium simulation, 300,000 snapshots separated by 10 ps each were taken as starting points for the nonequilibrium MD runs with the photoswitch switched into *trans*. Simulation times varied between 4 ps and 100 ns such that the number of samples in each time bin on a logarithmic time axis was roughly the same. That is, we ran 150,000 trajectories for 4 ps, 75,000 trajectories for 8 ps, and so on, up to 73 trajectories for 8 ns. In addition, 120 trajectories for a full 100 ns were collected. The total simulation time of these nonequilibrium trajectories amounts to ~21 μs and took about 5 months on a 96-core computer cluster. Both Figs. 4 and 5 are calculated from that full set of nonequilibrium trajectories.

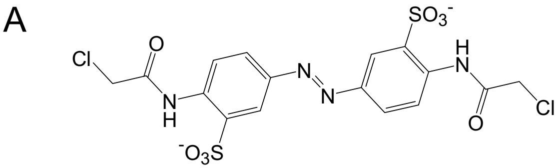
To calculate the water density on the protein surface (Fig. 5 and Movie S1), the 300,000 starting points, which resemble a *cis*-equilibrium ensemble, were aligned on each other by minimizing the rmsd of the C_α atoms from ordered secondary structure motives. An averaged structure was calculated from these starting points by averaging atom positions, which served as a reference to which all subsequent nonequilibrium structures were aligned together with their surrounding water molecules. The positions of the water-oxygens were then binned into cubes of 1 $Å^3$.

The residence times of water molecules were estimated by computing the average lifetime of individual water/protein hydrogen bonds along a 600-ns equilibrium MD trajectory with the photoswitch *in trans*. Time series of presence/rupture of individual hydrogen bonds were obtained by Gromacs with default thresholds of 0.35 nm for the donor–acceptor distance and 30°

for the acceptor–donor–hydrogen angle. Seven hydrogen bond donor or acceptor groups within the binding groove and 13 on the outside surface were used. Averaging over these groups resulted in hydrogen bond lifetimes of 40 ± 20 ps within the binding groove and 15 ± 5 ps on the outside surface. Neglecting transient hydrogen bond ruptures up to 10 ps yielded residence times of 60 ± 50 ps within the binding groove and 20 ± 10 ps on the outside surface.

The implicit solvent simulations were carried out according to a similar protocol with Newtonian dynamics at 300 K, using a generalized Born formalism (34). A *cis* structure already equilibrated in explicit solvent was further equilibrated for 300 ns in implicit solvent. A 700-ns production equilibrium run was then used to collect an ensemble of initial structures every 10 ns for the subsequent nonequilibrium simulations.

1. Burns DC, Zhang F, Woolley GA (2007) Synthesis of 3,3′-bis(sulfonato)-4,4′-bis(chloroacetamido)azobenzene and cysteine cross-linking for photo-control of protein conformation and activity. *Nat Protoc* 2(2):251–258.
2. Neri D, Szyperski T, Otting G, Senn H, Wüthrich K (1989) Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional 13C labeling. *Biochemistry* 28(19):7510–7516.
3. Live DH, Davis DG, Agosta WC, Cowburn D (1984) Observation of 1000-fold enhancement of nitrogen-15 NMR via proton-detected multiquantum coherences: Studies of large peptides. *J Am Chem Soc* 106(20):6104–6105.
4. Kay LE, Keifer P, Saarinen T (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J Am Chem Soc* 114(26):10663–10665.
5. Grzesiek S, Bax A (1992) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance nmr. *J Am Chem Soc* 114:6291–6293.
6. Grzesiek S, Bax A (1992) An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J Magn Reson* 99:201–207.
7. Kay LE, Ikura M, Tschudin R, Bax A (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J Magn Reson* 89:496–514.
8. Kay LE, Ikura M, Bax A (1990) Proton-proton correlation via carbon-carbon couplings: A three-dimensional NMR approach for the assignment of aliphatic resonances in proteins labeled with carbon-13. *J Am Chem Soc* 112:888–889.
9. Ikura M, Kay LE, Bax A (1991) Improved three-dimensional 1H-13C-1H correlation spectroscopy of a 13C-labeled protein using constant-time evolution. *J Biomol NMR* 1(3):299–304.
10. Grzesiek S, Anglister J, Bax A (1993) Correlation of backbone amide and aliphatic side-chain resonances in 13C/15N-enriched proteins by isotropic mixing of 13C magnetization. *J Magn Reson* 101(1):114–119.
11. Yamazaki T, Forman-Kay JD, Kay LE (1993) 2-dimensional NMR experiments for correlating C-13-beta and H-1-delta/epsilon chemical-shifts of aromatic residues in C-13-labeled proteins via scalar couplings. *J Am Chem Soc* 115:11054–11055.
12. Keller RL (2004) *The Computer Aided Resonance Tutorial* (Cantina).
13. Marion D, Kay LE, Sparks SW, Torchia DA, Bax A (1989) Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. *J Am Chem Soc* 111(4):1515–1517.
14. Ottiger M, Delaglio F, Bax A (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J Magn Reson* 131(2):373–378.
15. Hansen MR, Hanson P, Pardi A (2000) Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. *Methods Enzymol* 317:220–240.
16. Rückert M, Otting G (2000) Alignment of biological macromolecules in novel nonionic liquid crystalline media for nmr experiments. *J Am Chem Soc* 122(32):7793–7797.
17. Herrmann T, Güntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol* 319(1):209–227.
18. Herrmann T, Güntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J Biomol NMR* 24(3):171–189.
19. Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* 44(4):213–223.
20. Güntert P, Mumenthaler C, Wüthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* 273(1):283–298.
21. Güntert P (2004) Automated NMR structure calculation with CYANA. *Methods Mol Biol* 278:353–378.
22. Spronk CAEM, Linge JP, Hilbers CW, Vuister GW (2002) Improving the quality of protein structures derived by NMR spectroscopy. *J Biomol NMR* 22(3):281–289.
23. Mackerell AD, Jr., Feig M, Brooks CL, 3rd (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* 25(11):1400–1415.
24. Mackerell AD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102:3586–3616.
25. Spörlein S, et al. (2002) Ultrafast spectroscopy reveals subnanosecond peptide conformational dynamics and validates molecular dynamics simulation. *Proc Natl Acad Sci USA* 99(12):7998–8002.
26. Carstens H (2004) Konformationsdynamik lichtschaltbarer peptide: Molekulardynamiksimulationen und datengetriebene modellbildung [Conformational dynamics of photoswitchable peptides: Molecular dynamics and data-driven model building]. PhD thesis (Ludwig Maximilians Universität München, Munich). Available at http://edoc.ub.uni-muenchen.de/2268. German.
27. Vriend G (1990) WHAT IF: A molecular modeling and drug design program. *J Mol Graph* 8(1):52–56, 29.
28. Hooft RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. *Nature* 381(6580):272.
29. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8(4):477–486.
30. Van Der Spoel D, et al. (2005) GROMACS: Fast, flexible, and free. *J Comput Chem* 26(16):1701–1718.
31. Nguyen PH, Stock G (2006) Nonequilibrium molecular dynamics simulation of a photoswitchable peptide. *Chem Phys* 323:36–44.
32. Ihalainen JA, et al. (2008) α-Helix folding in the presence of structural constraints. *Proc Natl Acad Sci USA* 105(28):9588–9593.
33. Kozlov G, Gehring K, Ekiel I (2000) Solution structure of the PDZ2 domain from human phosphatase hPTP1E and its interactions with C-terminal peptides from the Fas receptor. *Biochemistry* 39(10):2572–2580.
34. Onufriev A, Bashford D, Case DA (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* 55(2):383–394.



B

MHHHHHHWDDDDLEVLFQ'G

```
          10         20         30         40         50
PKPGDIFEVE LAKNDNSLGI CVTGGVNTSV RHGGIYVKAV IPQGAAESDG
          60         70         80         90
RIHKGDRVLA VNGVSLEGAT HKQAVCTLRN TGQVVHLLLE KGQSPT
```

Fig. S1. Cross-linker and anchor sites in the protein sequence. (*A*) Photoswitch. This water-soluble thiol-reactive azobenzene derivative 3,3′-bis(sulfonato)-4,4′-bis(chloroacetamido)azobenzene (BSBCA) (1) was cross-linked to two cysteines in PDZ2. (*B*) Amino acid sequence of the second PDZ domain in human tyrosine-phosphatase 1E (hPTP1E). A His tag with a HRV 3C cleavage site (blue) was used for purification and cleaved before measurements. The cleavage site is marked with an apostrophe and the two Cys (S21C and E76C) that served as anchor points for the photoswitch are shown in red.

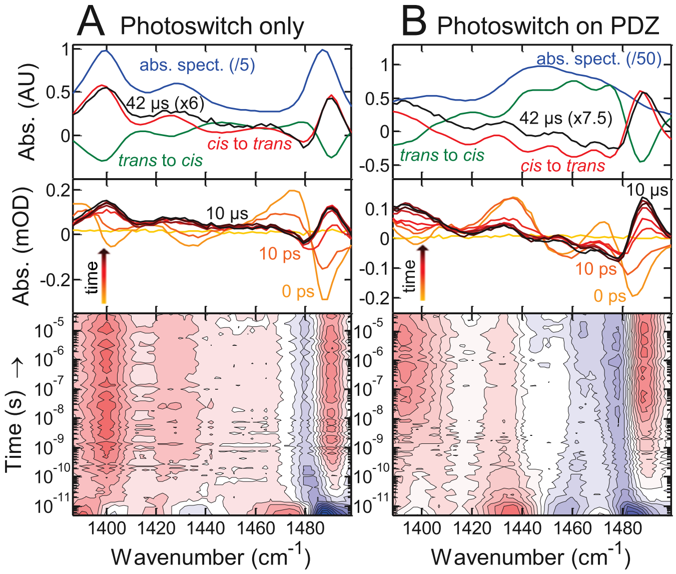


Fig. S2. IR spectra of the ring-mode (~1,390 cm^{-1}) and Amide II (~1,490 cm^{-1}) bands from the photoswitch. (*A* and *B*) (*Top*) Absolute (blue) and difference FTIR spectra (red and green) compared with the transient spectrum at 42 μs (black). (*Middle*) Transient spectra, at −1 ns (yellow), 0 s (light orange), and from 10 ps to 10 μs by decade (orange to black). (*Bottom*) Contour plot of the IR response. (*A*) Photoswitch alone; (*B*) photoswitch on PDZ2.

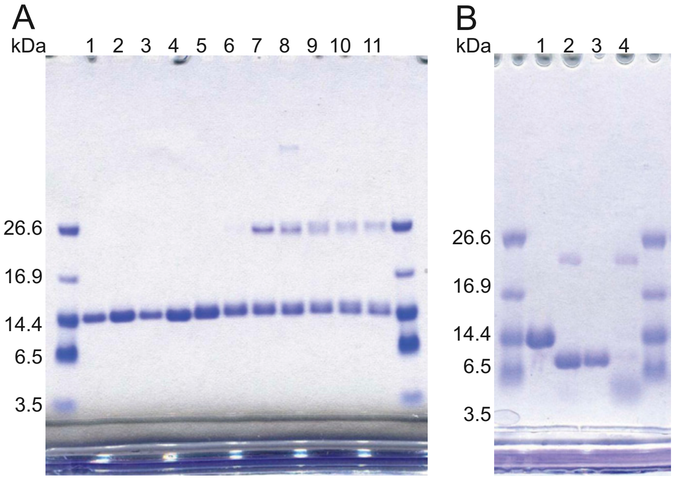


Fig. S3. SDS/PAGE analysis of linking the photoswitch (BSBCA) to PDZ2. Shown are 12% Bis-Tris gels (Novex NuPAGE; Invitrogen) with polypeptide SDS/PAGE molecular weight standards (Bio-Rad) Coomassie stained with Rotiphorese Blue R (Carl Roth GmbH). (*A*) After purification with a MonoQ anion exchange column (GE Healthcare Life Sciences). Lanes 1–5: PDZ2 with the photoswitch correctly linked to the two Cys (mass verified by ESI mass spectrometry). Lanes 6–11: Mixtures of incorrectly linked PDZ2 (monomers with two linkers, dimers, trimers). (*B*) Cleaving of the His tag. Lane 1, before cleaving; lane 2, after cleaving; lane 3, after cleaving and purification with HisTrap column (GE Healthcare Life Sciences); lane 4, HRV 3C protease and free His tag eluted from the HisTrap column.

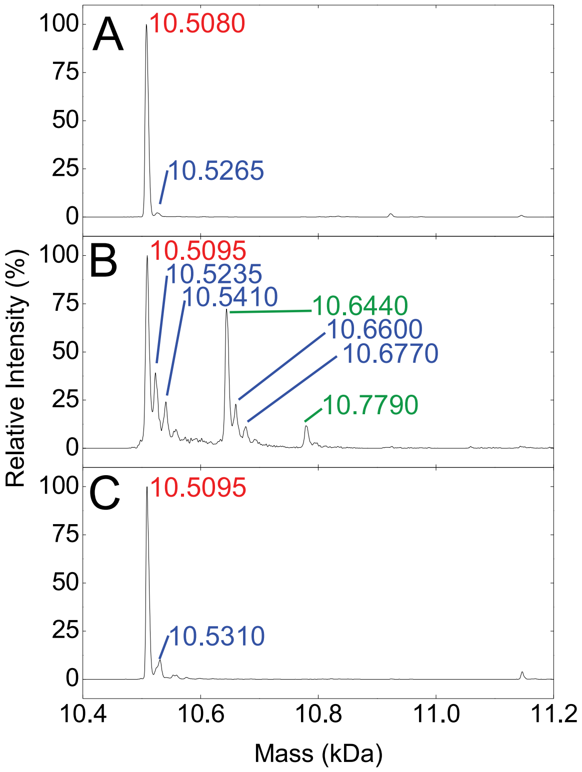


Fig. S4. Mass spectra of PDZ2 linked to the photoswitch. The peak corresponding to the expected mass is labeled in red, peaks originating from the oxidized sample are in blue, and peaks with an additional modification (+135 Da) are in green. (*A*) Before a time-resolved IR measurement, in 50 mM Tris·HCl, 150 mM NaCl, pH 8.5. (*B*) After a time-resolved IR measurement, in 50 mM Tris·HCl, 150 mM NaCl, pH 8.5. (*C*) After a time-resolved IR measurement, in 50 mM borate buffer, 150 mM NaCl, pH 8.5, and under exclusion of oxygen.

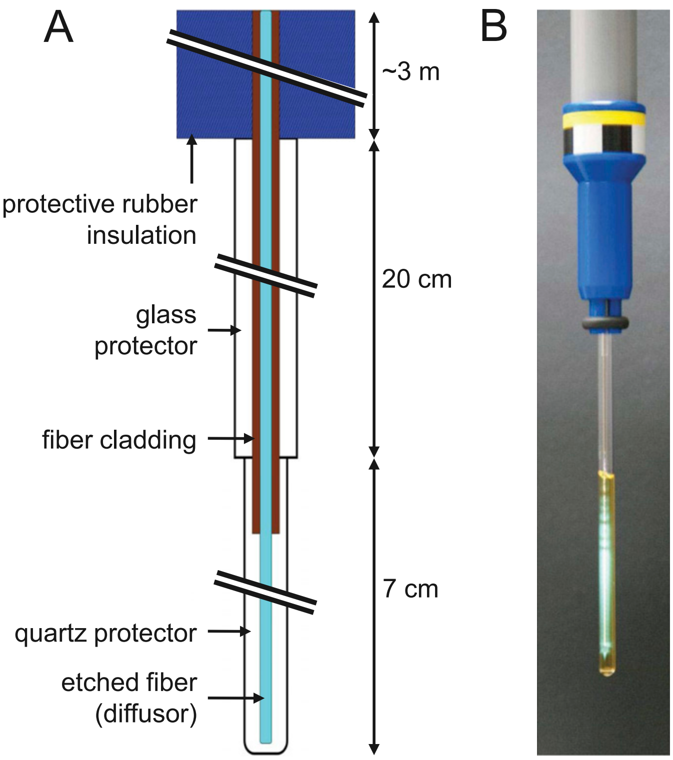


Fig. S5. Illumination of NMR sample. (*A*) Sketch of the NMR inline radial illumination probe (Polymicro Technologies). (*B*) Illumination of a sample in an NMR tube.

Table S1. Complete set of fit parameters for the black lines in Fig. 3

Probe wavenumber	a_0	a_1	τ_1	a_2	τ_2	β	a_3	τ_3
1,491 cm^{-1}	0.12	−0.16	0.015	−0.10	7.3	0.49	0	—
1,640 cm^{-1}	−0.07	0.0	0.015	0.09	7.3	0.49	−0.07	20,000

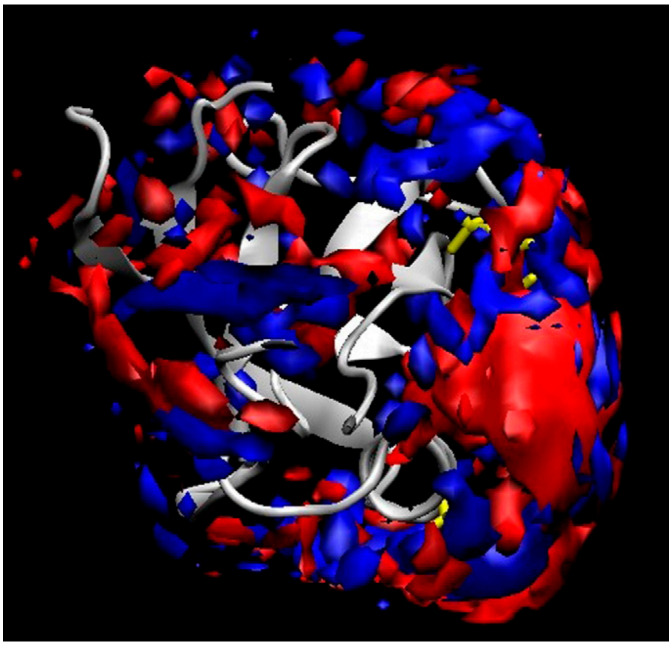
Amplitudes are in units of mOD (optical density), and time constants are in nanoseconds.

Table S2. Restraints for NMR structure calculation

Type	*cis*	*trans*
NOE	877	992
Photoswitch	2	29
Dihedral restraints	146	145
HN-RDCs		
Pf1	78	80
C12E5	76	76

Table S3. NMR structure statistics

Structural parameters of 20 NMR conformers	*cis*	*trans*
Pairwise cartesian rmsd, Å		
Global backbone heavy atoms	1.02	1.15
Global all heavy atoms	1.39	1.54
Ordered backbone heavy atoms	0.35	0.33
Ordered all heavy atoms	0.92	1.04
Ramachandran quality parameters, %		
Residues in most favored regions	86.7	86.3
Residues in allowed regions	11.2	8.8
Residues in additionally allowed regions	1.8	4.6
Residues in disallowed regions	0.3	0.3
Average rms deviation from current reliable structures, rms *Z*-scores		
Bond lengths	1.050	1.126
Bond angles	1.155	1.187
Omega angle restraints	1.303	1.486
Side-chain planarity	0.447	0.676
Improper dihedral distribution	0.908	0.983
Inside/outside distribution	0.951	0.986
Average deviation from current reliable structures, *Z*-scores		
First-generation packing quality	−1.833	−2.479
Ramachandran plot appearance	−2.014	−2.217
Chi-1/Chi-2 rotamer normality	−3.735	−3.859
Backbone conformation	−0.916	−0.916



Movie S1. Change of water density after photo-switching the azobenzene moiety that cross-links the binding groove of the PDZ2 domain. Red depicts increased water density; blue depicts decreased water density. The protein is shown by a gray ribbon; the photoswitch (visible only in part) is shown in yellow. The perturbation of the water network on the protein surface propagates within 100 ns around the protein. See Fig. 5 for details.

Movie S1

Chapter 5

Does bromodomain flexibility influence histone recognition?

Steiner, S., Magno, A., Huang, D. and Caflisch, A.; *FEBS letters*, 587(14): 2158-2163, 2013

FEBS Letters 587 (2013) 2158–2163

FEBS Letters

journal homepage: www.FEBSLetters.org

Does bromodomain flexibility influence histone recognition?

Sandra Steiner, Andrea Magno, Danzhi Huang, Amedeo Caflisch *

Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

ARTICLE INFO

Article history:
Received 27 February 2013
Revised 10 May 2013
Accepted 13 May 2013
Available online 25 May 2013

Edited by Robert B. Russell

Keywords:
Molecular dynamics
Post-translational modification
Histone
Epigenetic target

ABSTRACT

Bromodomains are protein modules that selectively recognize histones by binding to acetylated lysines. Here, we have carried out multiple molecular dynamics simulations of 20 human bromodomains to investigate the flexibility of their binding site. Some bromodomains show alternative side chain orientations of three evolutionarily conserved residues: the Asn involved in acetyl-lysine binding and two conserved aromatic residues. Furthermore, for the BAZ2B and CREBBP bromodomains we observe occlusion of the binding site which is coupled to the displacement of the two aromatic residues. In contrast to available structures, the simulations reveal large variability of the binding site accessibility. The simulations suggest that the flexibility of the bromodomain binding site and presence of self-occluded metastable states influence the recognition of acetyl-lysine on histone tails.

© 2013 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

1. Introduction

Bromodomains, protein modules of about 110 residues, recognize acetylated lysine side chains mainly in histones and are thus involved in transcriptional regulation [1,2]. In the human genome there are 46 proteins with a total of 61 different bromodomains, with up to six bromodomains per protein [3]. Due to the potential role of bromodomains in tumors and inflammation [4–6], large scale structural studies have been carried out with the ultimate goal to facilitate the discovery of small-molecule inhibitors able to interfere in the process of reading acetylated lysine [7]. Since 1999, when the first three-dimensional structure of a human bromodomain was solved [8], the crystal and/or solution structures of more than 40 human bromodomains have been reported [9,10]. All available structures show a conserved four-helix bundle topology (Fig. 1) in which the ZA-loop and BC-loop connect the first two α helices (called Z and A) and last two α helices (called B and C), respectively [10,11]. The acetyl-lysine binding site is very similar in all structures of bromodomains [10]. The BC-loop contributes the evolutionary conserved Asn side chain [12] which acts as hydrogen bond donor to the acetylated lysine side chain. Moreover, two conserved Tyr residues [13] are present in the ZA-loop and BC-loop, respectively (Fig. S1). In striking contrast to the abundance of available three-dimensional structures, there are only few computational studies on bromodomains [14–16]. In particular, it seems that the dynamic properties of human apo bromodomains have not been investigated yet by atomistic simulations. Here, we have studied the binding site flexibility of 20 bromodomains (covering seven of the eight families of human bromodomains) by explicit solvent MD simulations. The simulation results reveal a surprisingly high heterogeneity of the plasticity and accessibility of the acetyl-lysine binding site.

2. Results and discussion

A sequence alignment of the 20 simulated bromodomains with indication of important structural features and investigated key residues is given in Fig. S1. Family memberships are listed in Table S-I. The time series of root mean square deviation (RMSD) of the C_α atoms show that the overall fold of the simulated bromodomains is stable over the 1-μs time scale and 310 K temperature of the MD runs (Fig. S2). Most of the fluctuations are localized in the ZA-loop, BC-loop, and the termini (Fig. S3). The ZA-loop and BC-loop are involved in binding the histone tail whereas the termini are located far away from the acetyl-lysine binding site so that their motion can be neglected. Importantly, we observe multiple events of reversible rotation of the side chains investigated in the present study along the 0.5- to 1-μs time scale of the MD runs (Figs. S4 and S5). These rotations take place on a time scale about one order of magnitude smaller than the length of the simulations so that the following analysis is not marred by statistical errors.

We first investigated the flexibility of the conserved Asn which is directly involved in acetyl-lysine binding (i.e., Asn1944 in BAZ2B, Asn1168 in CREBBP, Asn1604 in TAF1(2), etc.). For most of the bromodomains studied here, there are many events of reversible

---

* Corresponding author. Fax: +41 (44) 635 68 62.
E-mail address: caflisch@bioc.uzh.ch (A. Caflisch).

0014-5793/$36.00 © 2013 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.febslet.2013.05.032

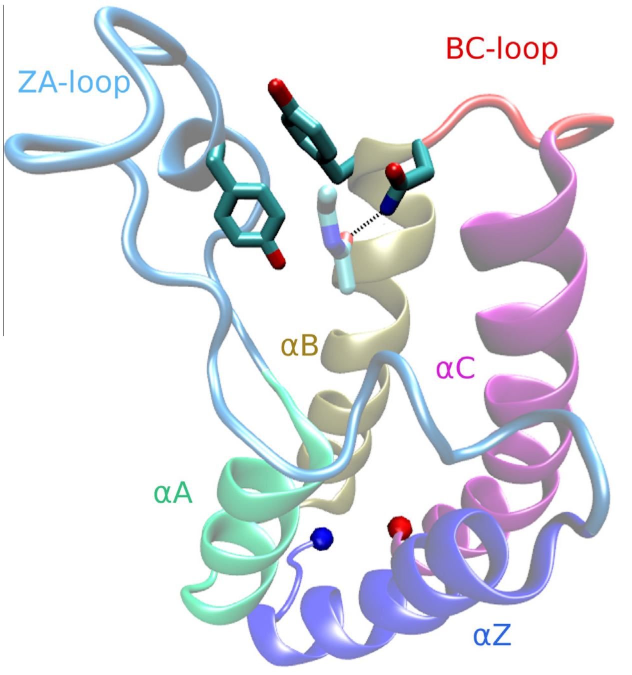


Fig. 1. Ribbon illustration of the crystal structure of the complex between CREBBP and acetyl-lysine (PDB code 3P1C). Each of the four α-helices and two binding site loops is displayed with a different color. The side chains of the conserved Tyr1125 of the ZA-loop, and Tyr1167-Asn1168 of the BC-loop are emphasized (sticks) together with the side chain of acetyl-lysine (sticks, light colors). The N-terminus and C-terminus are shown with a blue and red sphere, respectively.

rotations of the side chain amide of the conserved Asn (Figs. S4 and S5). The rotation (of about 180 degrees) around the χ_2 angle of the conserved Asn switches between an orientation in which the side chain –NH$_2$ group caps the C-terminal turn of α helix B (by donating a hydrogen bond to the backbone carbonyl of the n-4 upstream residue, e.g., the hydrogen bond between the Asn1168 side chain and the carbonyl group of residue 1164 in CREBBP) to a completely solvent exposed orientation of the –NH$_2$ group (Fig. 2, top). Note that the former orientation is observed in most (13 of 17) X-ray structures of the apo bromodomains used in this study (three of them have Thr (PHIP(2) and WDR9(2)) or Tyr (ASH1L) instead of the conserved Asn), as well as in the X-ray structures of BRD4(1) with three different diacetylated peptides (PDB codes 3UVW, 3UVY, and 3UW9) and in the CREBBP/acetyl-lysine complex (PDB code 3P1C). Interestingly, in seven of the 17 bromodomains the Asn –NH$_2$ group is oriented towards the solvent during more than one third of the simulation time (Figs. S4 and S5). These bromodomains are: TAF1L(2) (90% of the snapshots with solvent-exposed orientation of the Asn –NH$_2$ group), TAF1(2) (81%), TAF1(1) (68%), GCN5L2 (44%), ATAD2 (43%), BAZ2B (39%), and PB1(2) (38%). For TAF1L(2), TAF1(2), and TAF1(1) (all belonging to a single clade of family VII in the bromodomain phylogenetic tree [10]) the solvent exposed orientation of the Asn -NH$_2$ group is stabilized by a water-bridged hydrogen bond between the Asn side chain carbonyl and the hydroxyl group of the Ser in position n-4 upstream (Fig. 2, bottom). It has to be noted that most bromodomains have Ala or Cys at position n-4 upstream of the conserved Asn so that these side chains cannot form a (water-bridged) hydrogen bond. Note also that an in-depth analysis of the water molecules in the

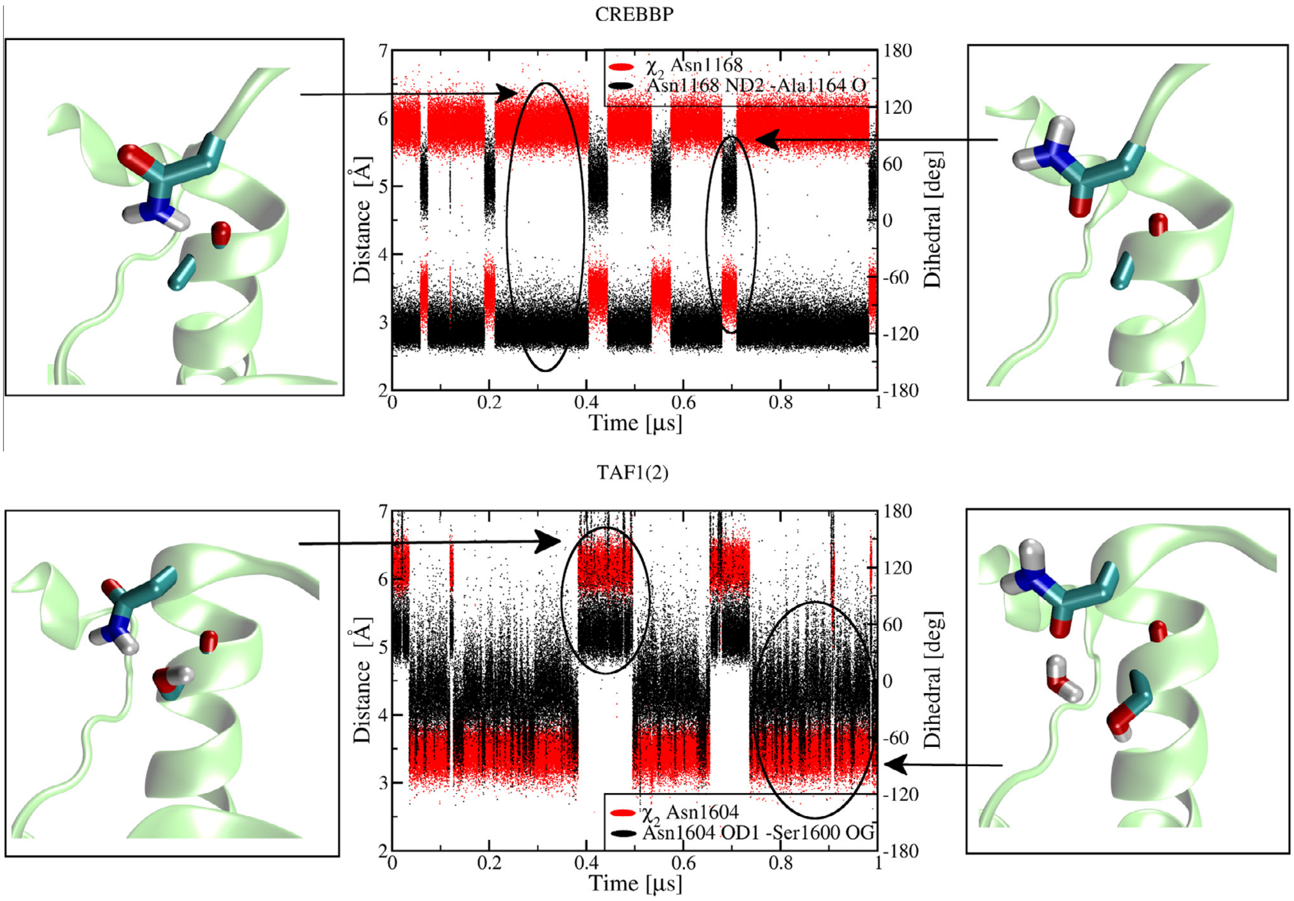


Fig. 2. Rotational flexibility of the conserved Asn side chain as observed in MD simulations of CREBBP (top) and TAF1(2) (bottom). (Middle) Time series of Asn χ_2 angle of conserved Asn (red) and distance between Asn side chain and residue n-4 upstream (black). (Left and right) Representative MD snapshots.

binding site along the MD trajectories goes beyond the scope of this paper and will be presented elsewhere together with a comparison with the five water molecules that seem to be conserved in the crystal structures [10].

We further analyzed the motion of residues adjacent to the conserved Asn to shed light onto the accessibility of the acetyl-lysine binding site. The MD trajectories of some of the 20 simulated bromodomains unmask a large flexibility of the side chain of the evolutionary conserved Tyr (Phe in BAZ2B and SMARCA4) in the BC-loop, i.e., the residue preceding in the sequence the Asn involved in acetyl-lysine binding (Figs. S4 and S5). The side chain of this Tyr (or Phe) rotates from an inward orientation in which the aromatic ring is part of the binding site surface (χ_1 angle in trans) to an outward orientation in which the same side chain points towards the exterior surface of the bromodomain (χ_1 angle values around −60 degrees) (Fig. 3). The five bromodomains with a population larger than 10% of the outward orientation of the BC-loop Tyr (or Phe) side chain are BRDT(1) (100%), CREBBP (31%), BAZ2B (24%), SMARCA4 (21%), and BRD1 (11%). This simulation result is a novel finding because only two of the more than 50 bromodomain crystal and solution structures (i.e., the X-ray structure of BRDT(1) and the NMR structure of KIAA1240, PDB codes 2RFJ and 2DKW, respectively) show a similar outward orientation of the side chain of the conserved Tyr on the BC-loop. Note that in contrast to the solution structure, the crystal structure of KIAA1240 (PDB code 3LXJ) does not show the outward orientation of the conserved Tyr.

To further investigate the influence of the displacement of the BC-loop Tyr on the binding site accessibility, two-dimensional histograms were calculated for each of the 20 simulated bromodomains (Figs. 3(b) and S6). According to the two-dimensional histograms the 20 bromodomains can be assigned to three different groups. The largest group consists of the bromodomains with an open binding site, and in which the side chain of the conserved Tyr of the BC-loop is part of the binding site surface (χ_1 angle always in the trans orientation, e.g., TAF1(2)). The second group includes the bromodomains in which the conserved Tyr changes its orientation without affecting the binding site accessibility (e.g., BRD1 and SMARCA4). The third group includes the bromodomains

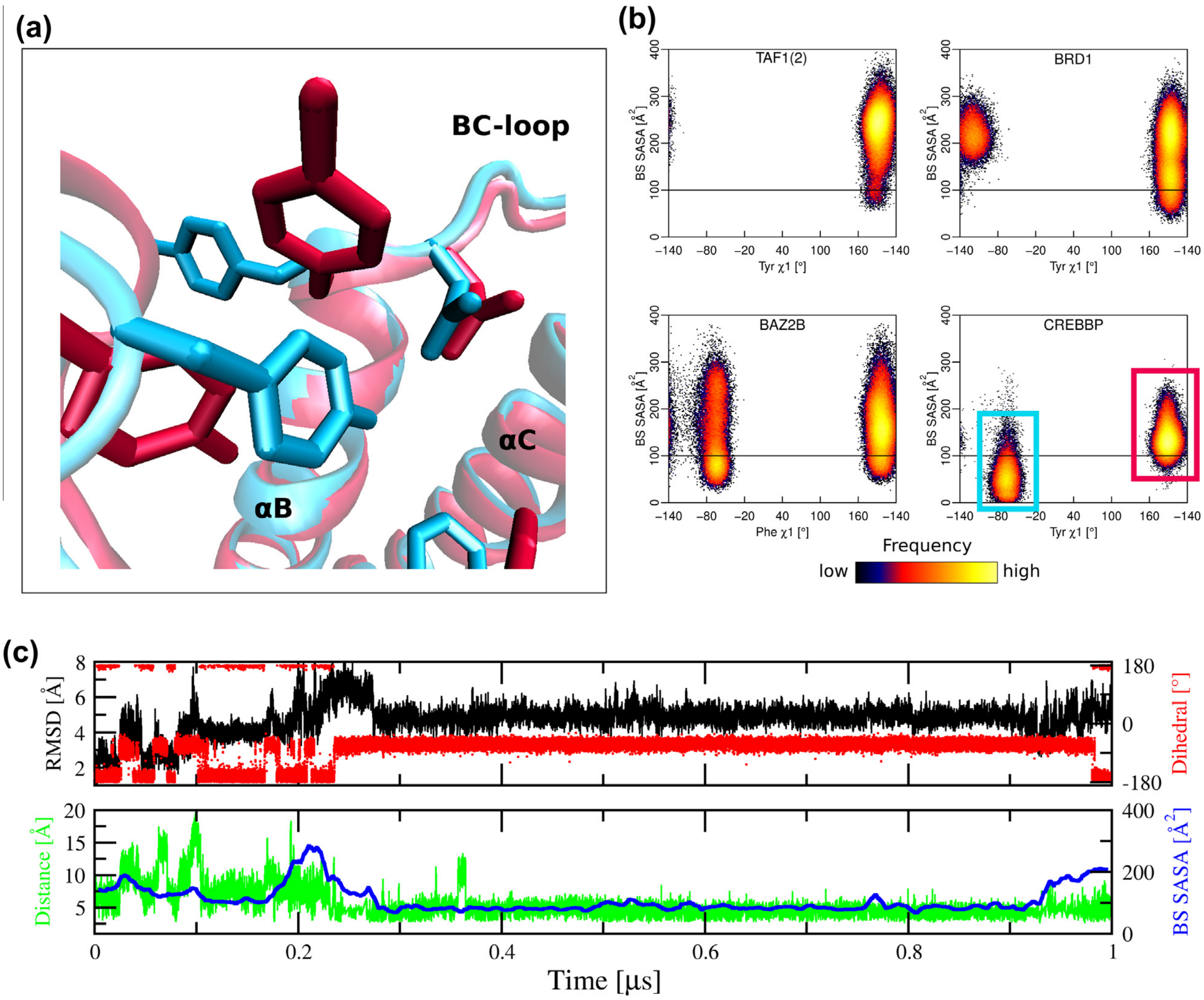


Fig. 3. Swapping out of the aromatic residue preceding the conserved Asn in the BC-loop. (a) Representative snapshots extracted from the MD simulations of CREBBP. (b) Two-dimensional histograms of binding site solvent accessible surface area (SASA) versus side chain χ_1 angle of the conserved Tyr (Phe in BAZ2B) in the BC-loop. The horizontal line at 100 Å^2 is an arbitrary threshold drown solely to help the eye. The cyan and red rectangles in the histogram of CREBBP emphasize the metastable states corresponding to the cyan and red structures, respectively, in panel (a). The black to yellow coloring of frequency values follows a logarithmic scale and is shown in the legend bar. (c) Time series of four geometrical variables used to analyze the binding site. This panel shows run 2 of BAZ2B while the equivalent time series for the other MD runs are shown in Figs. S7–S10. The variables are the RMSD of the C$_\alpha$ atoms of the ZA-loop from the crystal structure (black), the Phe1943 χ_1 dihedral angle (red), the binding site SASA (blue, running average calculated over time intervals of 10 ns for clarity), and the distance between the hydroxyl oxygen of Tyr1901 and the nitrogen atom of the Asn1944 side chain (green).

in which the outward orientation of the BC-loop Tyr promotes a rearrangement of other binding site residues resulting in a partially occluded, i.e., shallower binding site (e.g., BAZ2B and CREBBP, Figs. 3(b) and 4). More precisely, one observes large fluctuations of the ZA-loop that are associated frequently with the swapping out of the BC-loop Tyr (Figs. 3(c), S7 and S8) and with changes in the accessibility of the binding site (Figs. S9 and S10). In particular, in the BAZ2B and CREBBP simulation segments during which the conserved Tyr of the BC-loop points outward, the conformational plasticity of the ZA-loop results (in 20% to 25% of the saved snapshots) in the formation of a direct hydrogen bond between Asn1168 and ZA-loop Tyr1125 in CREBBP (Figs. 3(c) and 4), and Asn1944 and ZA-loop Tyr1901 in BAZ2B. Note that these two side chains are separated by two water molecules in the X-ray structures (PDB codes 3DWY and 3G0L). In essence, it emerges from the MD trajectories that the ZA-loop Tyr side chain can occupy the same position and can be stabilized by the same hydrogen bond to Asn as the natural ligand acetyl-lysine (Fig. 4).

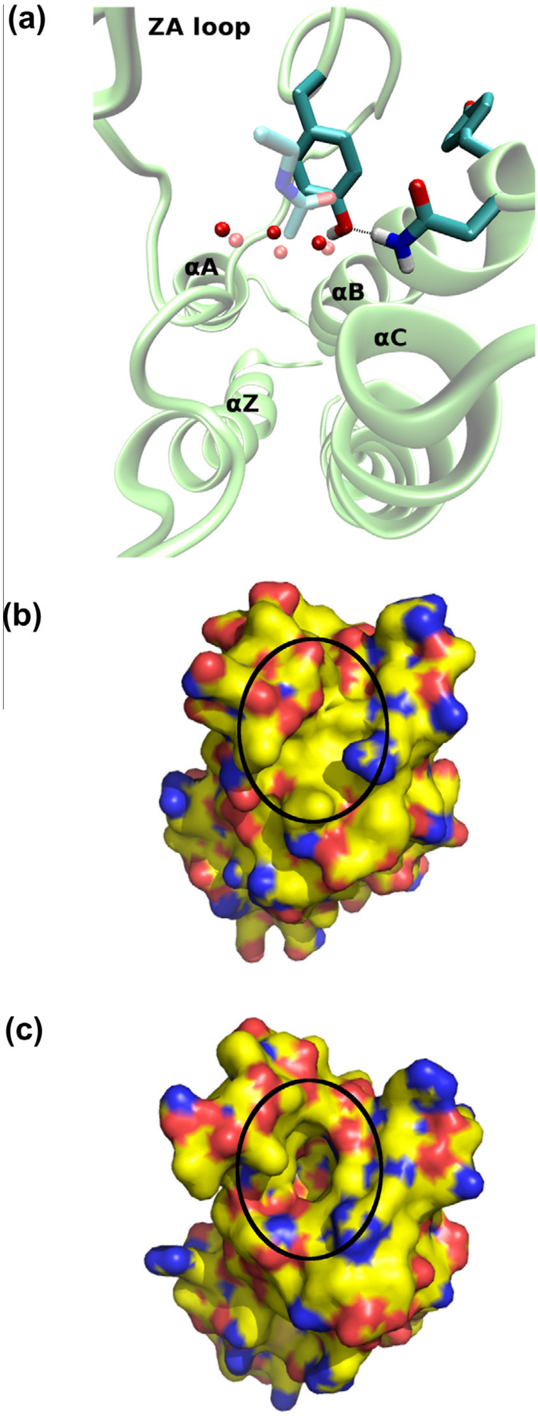


Fig. 4. Occlusion of the acetyl-lysine binding site by the conserved Tyr of the ZA-loop. (a) The ribbon structure is a representative MD snapshot of the CREBBP simulations in which the conserved Tyr of the ZA-loop is involved in a hydrogen bond with the Asn that binds acetyl-lysine. The three conserved side chains (sticks) are shown together with three water molecules (red spheres). The position of acetyl-lysine (light-colored sticks) and three water molecules (light red spheres) from the crystal structure of the CREBBP/acetyl-lysine complex (PDB code 3P1C) was obtained by structural superposition of the C_α atoms. Surface representation of a representative MD snapshot with partially occluded, i.e., shallower, binding site (b) and the crystal structure of CREBBP (PDB code 3P1C) (c). The color coding reflects atomic elements (carbon, nitrogen, and oxygen atoms in yellow, blue, and red, respectively). The location of the acetyl-lysine binding site is indicated with a black oval.

The ZA-loop shows significant plasticity and large deviations from the crystal structure (Figs. 3(c), S7 and S8). This flexibility is consistent with the structural heterogeneity of the ZA-loop observed by overlapping the available X-ray structures [10] as well as with the length of the ZA-loop (about 25–30 residues, Fig. S1). Figs. 3(c), S9 and S10 show that the displacement of the ZA-loop influences the binding site accessibility (note that 4 of the 6 residues used to calculate the binding site accessibility are in the ZA-loop). In particular, the accessibility can change significantly in the MD trajectory segments during which the ZA-loop RMSD from the crystal structure is large, while the accessibility is less variable in segments with mainly rigid ZA-loop. Interestingly, the ZA-loop conformational flexibility has been shown to play an important role at the protein-protein interface for HIV-1 Tat binding to bromodomain PCAF [14]. In the present simulations of CREBBP and BAZ2B, the rearrangement of the ZA-loop and thus of several residues lining the binding site reduces its aperture which could hinder histone peptide recognition.

In conclusion, the MD trajectories reveal a heterogeneous flexibility of the acetyl-lysine binding site in different bromodomains. The analysis of the trajectories shows that the accessibility of the binding site ranges from fully preserved (for the majority of bromodomains) to almost completely inaccessible during simulation segments of hundreds of ns (e.g., in BAZ2B and CREBBP). Of practical importance, the presence of metastable states in a small subset of bromodomains can be exploited [17–19] to design selective small-molecule inhibitors few of which have been reported as of today [4,20–23]. The present simulation results shed some light on the binding mechanism of histone tails to bromodomains which seems to be mainly conformational selection, though a (partial) induced fit cannot be definitively excluded. More precisely, with the present simulations of apo bromodomains it is not possible to exclude a ligand-promoted displacement of some of the side chains in the acetyl-lysine binding site. The self-occluded conformation of the acetyl-lysine binding groove suggests a possible auto-regulatory mechanism of histone tail recognition and binding.

3. Methods

The coordinates of the 20 bromodomains were downloaded from the protein database [24]. To reproduce neutral pH conditions the side chains of aspartates and glutamates were negatively charged, those of lysines and arginines were positively charged, and the histidine side chains were neutral. For each bromodomain the crystal water molecules close to the binding site and with low B factor value were kept while the remaining water molecules were deleted. For most bromodomains a B-factor of 15 Å^2 was used as threshold for keeping crystallographic water molecules. For the few bromodomains with overall higher B-factors a threshold of 30 Å^2 was used. With this arbitrary choice of thresholds between

4 and 15 crystal water molecules were retained for each bromodomain. Importantly, in our simulations the structured water molecules exchange rapidly (residence times ranging from 0.01 to 1 ns) with the water molecules in the bulk solvent. Furthermore, in test simulations without crystal water molecules, bulk water molecules rapidly occupied the conserved positions in the binding site. Thus, the exact number of kept crystal waters does not influence the results of our simulations because individual runs have a length of 0.5 or 1 μs.

Subsequently, the structure was solvated in a water box whose size was chosen to have a minimal distance of 11 Å between the boundary and any atom of the protein. Solvation water molecules within 2.4 Å of any heavy atom of the protein or crystal water were removed. The simulation system contained sodium and chloride ions to approximate an ionic strength of about 150 mM and to compensate for the total charge of each bromodomain. The MD simulations were carried out with Gromacs 4.5.4 [25] using the CHARMM PARAM22 force field [26] and the TIP3P model of water [27].

Periodic boundary conditions were applied and electrostatic interactions were evaluated using the particle-mesh Ewald summation method [28]. The van der Waals interactions were truncated at a cutoff of 10 Å. The temperature of 310 K was kept constant by an external bath with velocity rescaling [29] and the pressure was kept close to 1 atm by the Berendsen barostat. The SHAKE algorithm was used to fix the covalent bonds involving hydrogen atoms. The integration time step was 2 fs and snapshots were saved every 10 ps and 4 ps for the 1-μs and 0.5-μs runs, respectively.

Multiple independent MD runs of 1 μs each were carried out for four bromodomains belonging to four different families: BAZ2B (family V), BRD1 (IV), CREBBP (III), and TAF1(2) (VII). In addition, 16 bromodomains covering all but one of the families were simulated for 0.5 μs each (Table S-I). The crystal structures solved by the Structural Genomics Consortium [10] were used as starting conformations.

The solvent accessible surface area (SASA) was calculated with the SURF tool of Wordom, using the GEPOL algorithm [30,31] and 1.4 Å as radius of the water probe. The following procedure was used to define the acetyl-lysine binding site for the SASA calculation. First, using the X-ray structure of the CREBBP/acetyl-lysine complex (PDB code 3P1C) residues were selected which had at least three heavy atoms within 6 Å of the ligand. From this set Tyr1167 and Asn1168 were excluded to avoid redundancy in the two variables used for the two-dimensional histograms. Of the six remaining residues (Pro1110, Phe1111, Val1115, Tyr1125, Ala1164, Val1174), only the side chain atoms were used to evaluate the binding site SASA. For the other bromodomains the equivalent residues upon structural overlap were chosen. Robustness of the selection of binding site residues is shown in Fig. S11.

Acknowledgments

The simulations were carried out on the Schrödinger cluster at the University of Zurich. This work was supported by a grant of the Swiss National Science Foundation to A.C.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ejor.2012.04.007.

References

[1] Zeng, L. and Zhou, M.-M. (2002) Bromodomain: an acetyl-lysine binding domain. FEBS Lett. 513 (1), 124–128.

[2] Filippakopoulos, P. and Knapp, S. (2012) The bromodomain interaction module. FEBS Lett. 586 (17), 2692–2704.

[3] Arrowsmith, C.H., Bountra, C., Fish, P.V., Lee, K. and Schapira, M. (2012) Epigenetic protein families: a new frontier for drug discovery. Nat. Rev. Drug Discov. 11 (5), 384–400.

[4] Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W.B., Fedorov, O., Morse, E.M., Keates, T., Hickman, T.T., Felletar, I., Philpott, M., Munro, S., McKeown, M.R., Wang, Y., Christie, A.L., West, N., Cameron, M.J., Schwartz, B., Heightman, T.D., Thangue, N.L., French, C.A., Wiest, O., Kung, A.L., Knapp, S. and Bradner, J.E. (December 2010) Selective inhibition of bet bromodomains. Nature 468 (7327), 1067–1073.

[5] Zuber, J., Shi, J., Wang, E., Rappaport, A.R., Herrmann, H., Sison, E.A., Magoon, D., Qi, J., Blatt, K., Wunderlich, M., Taylor, M.J., Johns, C., Chicas, A., Mulloy, J.C., Kogan, S.C., Brown, P., Valent, P., Bradner, J.E., Lowe, S.W. and Vakoc, C.R. (2011) Rnai screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. Nature 478 (7370), 524–528.

[6] Dawson, M.A., Kouzarides, T. and Huntly, B.J. (2012) Targeting epigenetic readers in cancer. N. Engl. J. Med. 367 (7), 647–657. PMID: 22894577.

[7] Hewings, D.S., Rooney, T.P.C., Jennings, L.E., Hay, D.A., Schofield, C.J., Brennan, P.E., Knapp, S. and Conway, S.J. (2012) Progress in the development and application of small molecule inhibitors of bromodomain-acetyl-lysine interactions. J. Med. Chem. 55 (22), 9393–9413.

[8] Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, A.K. and Zhou, M.-M. (1999) Structure and ligand of a histone acetyltransferase bromodomain. Nature 399 (6735), 491–496.

[9] Mujtaba, S., Zeng, L. and Zhou, M. (2007) Structure and acetyl-lysine recognition of the bromodomain. Oncogene 26 (37), 5521–5527.

[10] Filippakopoulos, P., Picaud, S., Mangos, M., Keates, T., Lambert, J.-P., Barsyte-Lovejoy, D., Felletar, I., Volkmer, R., Müller, S., Pawson, T., Gingras, A.-C., Arrowsmith, C.H. and Knapp, S. (2012) Histone recognition and large-scale structural analysis of the human bromodomain family. Cell 149 (1), 214–231.

[11] Vidler, L.R., Brown, N., Knapp, S. and Hoelder, S. (2012) Druggability analysis and structural classification of bromodomain acetyl-lysine binding sites. J. Med. Chem. 55 (17), 7346–7359.

[12] Owen, D.J., Ornaghi, P., Yang, J.-C., Lowe, N., Evans, P.R., Ballario, P., Neuhaus, D., Filetici, P. and Travers, A.A. (2000) The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase Gcn5p. EMBO J. 19 (22), 6141–6149.

[13] Furdas, S.D., Carlino, L., Sippl, W. and Jung, M. (2012) Inhibition of bromodomain-mediated protein–protein interactions as a novel therapeutic strategy. Med. Chem. Commun. 3, 123–134.

[14] Pantano, S., Marcello, A., Ferrari, A., Gaudiosi, D., Sabò, A., Pellegrini, V., Beltram, F., Giacca, M. and Carloni, P. (2006) Insights on HIV-1 Tat:P/CAF bromodomain molecular recognition from in vivo experiments and molecular dynamics simulations. Proteins: Struct. Funct. Bioinform. 62 (4), 1062–1073.

[15] Pizzitutti, F., Giansanti, A., Ballario, P., Ornaghi, P., Torreri, P., Ciccotti, G. and Filetici, P. (2006) The role of loop ZA and Pro371 in the function of yeast Gcn5p bromodomain revealed through molecular dynamics and experiment. J. Mol. Recognit. 19 (1), :1–9.

[16] Eichenbaum, K.D., Rodríguez, Y., Mezei, M. and Osman, R. (2010) The energetics of the acetylation switch in p53-mediated transcriptional activation. Proteins: Struct. Funct. Bioinform. 78 (2), 447–456.

[17] Ekonomiuk, D., Su, X.-C., Ozawa, K., Bodenreider, C., Lim, S.P., Otting, G., Huang, D. and Caflisch, A. (August 2009) Flaviviral protease inhibitors identified by fragment-based library docking into a structure generated by molecular dynamics. J. Med. Chem. 52 (15), 4860–4868.

[18] Zhao, H., Huang, D. and Caflisch, A. (September 2012) Discovery of tyrosine kinase inhibitors by docking into an inactive kinase conformation generated by molecular dynamics. ChemMedChem 7, 1983–1990.

[19] Laine, E., Martnez, L., Ladant, D., Malliavin, T. and Blondel, A. (August 2012) Molecular motions as a drug target: mechanistic simulations of anthrax toxin edema factor function led to the discovery of novel allosteric inhibitors. Toxins (Basel) 4 (8), 580–604.

[20] Nicodeme, E., Jeffrey, K.L., Schaefer, U., Beinke, S., Dewell, S., Chung, C.-W., Chandwani, R., Marazzi, I., Wilson, P., Coste, H., White, J., Kirilovsky, J., Rice, C.M., Lora, J.M., Prinjha, R.K., Lee, K. and Tarakhovsky, A. (December 2010) Suppression of inflammation by a synthetic histone mimic. Nature 468 (7327), 1119–1123.

[21] Gerona-Navarro, G., Yoel-Rodrguez, Mujtaba, S., Frasca, A., Patel, J., Zeng, L., Plotnikov, A.N., Osman, R. and Zhou, M.-M. (February 2011) Rational design of cyclic peptide modulators of the transcriptional coactivator CBP: a new class of p53 inhibitors. J. Am. Chem. Soc. 133 (7), :2040–2043.

[22] Chung, C.-W., Dean, A.W., Woolven, J.M. and Bamborough, P. (January 2012) Fragment-based discovery of bromodomain inhibitors part 1: Inhibitor binding modes and implications for lead discovery. J. Med. Chem. 55 (2), 576–586.

[23] Bamborough, P., Diallo, H., Goodacre, J.D., Gordon, L., Lewis, A., Seal, J.T., Wilson, D.M., Woodrow, M.D. and Chung, C.-W. (January 2012) Fragment-based discovery of bromodomain inhibitors part 2: Optimization of phenylisoxazole sulfonamides. J. Med. Chem. 55 (2), 587–596.

[24] Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1978) The protein data

bank: a computer-based archival file for macromolecular structures. Arch. Biochem. Biophys. 185 (2), 584–591.

[25] Spoel, D.V.D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E. and Berendsen, H.J.C. (December 2005) Gromacs: fast, flexible, and free. J. Comput. Chem. 26 (16), 1701–1718.

[26] MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L.J., et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B 102, 35863616.

[27] Jorgensen, W.L., Chandrasekhar, J., Madura, J., Impey, R.W. and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926935.

[28] Darden, T., York, D. and Pedersen, L.G. (1993) Particle mesh Ewald: an $n\log(n)$ method for Ewald sums in large systems. J. Chem. Phys. 98, 10089.

[29] Bussi, G., Donadio, D. and Parrinello, M. (2007) Canonical sampling through velocity rescaling. J. Chem. Phys. 126 (1), 014101.

[30] Seeber, M., Cecchini, M., Rao, F., Settanni, G. and Caflisch, A. (October 2007) Wordom: a program for efficient analysis of molecular dynamics simulations. Bioinformatics 23 (19), 2625–2627.

[31] Seeber, M., Felline, A., Raimondi, F., Muff, S., Friedman, R., Rao, F., Caflisch, A. and Fanelli, F. (April 2011) Wordom: a user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. J. Comput. Chem. 32 (6), 1183–1194.

Does bromodomain flexibility influence histone recognition?

SUPPLEMENTARY MATERIAL

S. Steiner[a], A. Magno[a], D. Huang[a], and A. Caflisch[a*]

[a]Department of Biochemistry
University of Zürich, Winterthurerstrasse 190
CH-8057 Zürich, Switzerland
Phone: (+41 44) 635 55 21, FAX: (+41 44) 635 68 62
email: caflisch@bioc.uzh.ch

[*]Corresponding author

keywords: molecular dynamics, post-translational modifications, histones, epigenetic targets

May 9, 2013

Simulation systems

Bromodomain	Family	PDB entry	Number of independent runs	Total sampling (μs)	Number of available X-ray/NMR structures
CECR2	I	3NXB	1	0.5	1
FALZ	I	3UV2	1	0.5	2
GCN5L2	I	3D7C	1	0.5	2
PCAF	I	3GG3	1	0.5	2
BRD4(1)	II	2OSS	1	0.5	1
BRDT(1)	II	2RFJ	1	0.5	1
BRD3(2)	II	2OO1	1	0.5	2
PHIP(2)	III	3MB3	1	0.5	1
WDR9(2)	III	3Q2E	1	0.5	1
CREBBP	III	3DWY	5	**4.5**	1
EP300	III	3I3J	1	0.5	1
ATAD2	IV	3DAI	1	0.5	1
BRD1	IV	3RCW	5	**4.5**	1
BAZ2B	V	3G0L	5	**4.5**	1
TAF1(1)	VII	3UV5	1	0.5	2
TAF1L(2)	VII	3HMH	1	0.5	1
TAF1(2)	VII	3UV4	4	**4.0**	2
ASH1L	VIII	3MQM	1	0.5	1
PB1(2)	VIII	3HMF	1	0.5	1
SMARCA4	VIII	3UVD	1	0.5	3

Table S-I: The digit in parentheses in the first column indicates the position of the bromodomain along the polypeptide sequence, e.g., TAF1(2) is the second bromodomain of the human Transcription initiation factor TFIID subunit 1 (TAF1). Multiple independent MD runs were carried out for the four bromodomains in boldface. The family classification and number of available X-ray/NMR structures are taken from Ref.[1]

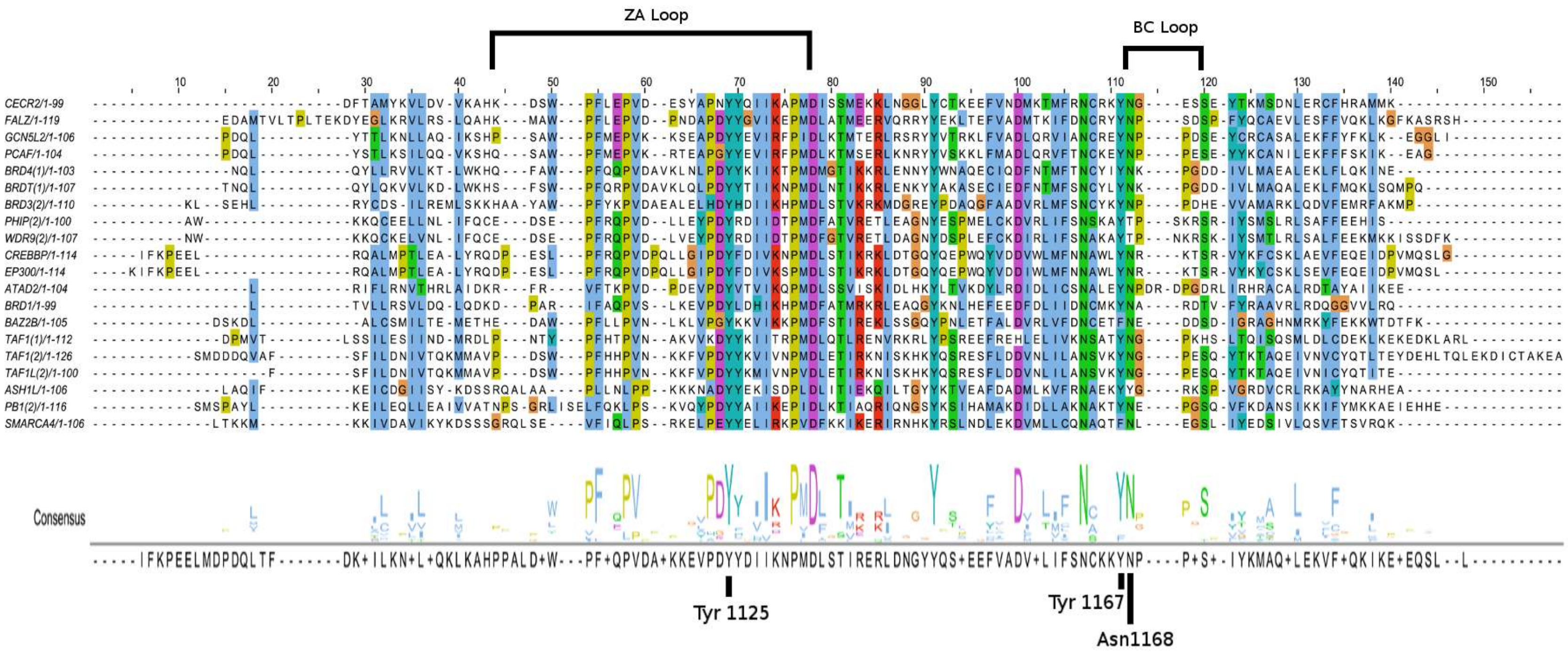


Figure S1: **Sequence alignment of the 20 simulated bromodomains.** The sequences were aligned first with MUSCLE ,[2] and subsequently realigned with ClustalW[3] using default setting parameters. Residues are colored following the ClustalX color scheme. The three residues in the bottom emphasize the evolutionary conserved Tyr side chain on the ZA-loop, i.e., Tyr1125 in CREBBP, Tyr599 in BRD1, Tyr1901 in BAZ2B, and Tyr1561 in TAF1(2), and the conserved aromatic residue and Asn at the beginning of the BC-loop, i.e., Tyr1167-Asn1168 in CREBBP, Tyr641-Asn642 in BRD1, Phe1943-Asn1944 in BAZ2B, and Tyr1603-Asn1604 in TAF1(2). The conserved residues are labeled only for CREBBP.

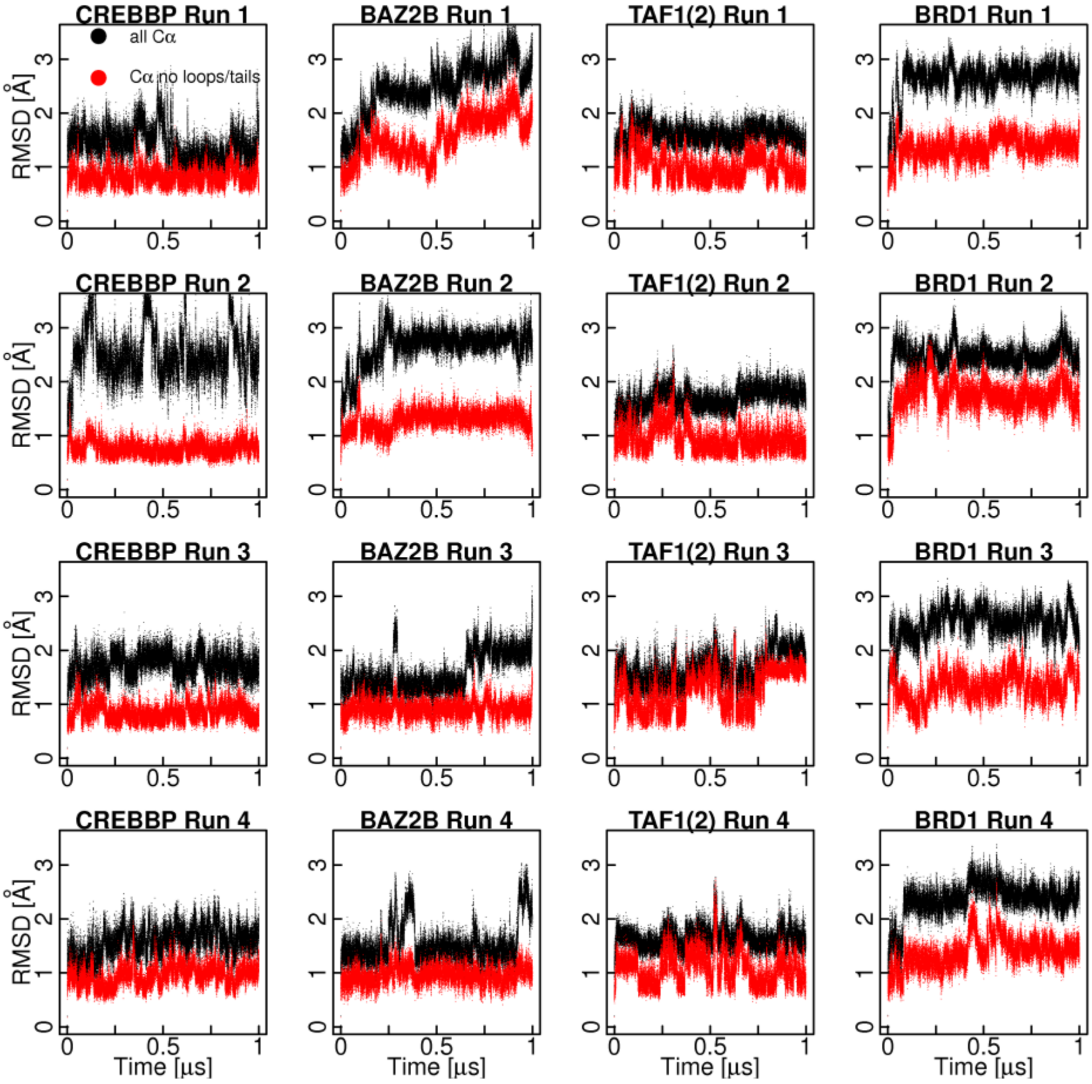


Figure S2: **Root mean square deviation (RMSD) of C_α atoms.** The time series of the RMSD from the minimized structures (i.e., starting structure of the MD runs) is plotted for all C_α atoms (black line) and excluding the C_α atoms in the ZA-loop, BC-loop, and the termini (red line). For CREBBP C_α's of residues 1104-1134 and 1168-1172 as well as 3 terminal residues on each N- and C-terminus were excluded for the latter calculation. For BAZ2B, BRD1 and TAF1(2) the corresponding residues upon structural overlap were neglected. Note that similar RMSD values are observed for the other 16 bromodomains.

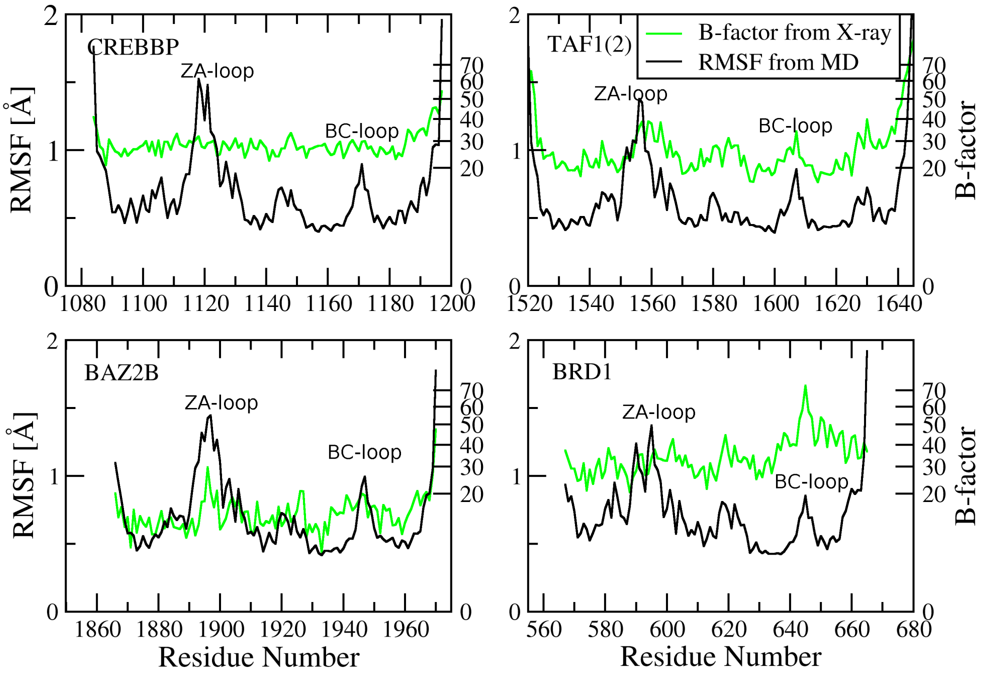


Figure S3: **Root mean square fluctuations (RMSF).** Profile of the RMSF of the C_α atoms (black line, y-axis on the left) of CREBBP, BAZ2B, BRD1 and TAF1(2) and crystallographic B-factors (green line, y-axis on the right) as a function of residue number. The RMSF values were averaged over simulation intervals of 5 ns. The fluctuations are highest for the ZA-loop, BC-loop, and termini. Note that the major discrepancies between RMSF along the simulations and crystallographic B-factors are due to crystal contacts in the latter, e.g., for the loops of CREBBP which are stabilized by the nearest neigbors in the crystal (PDB code 3DWY).

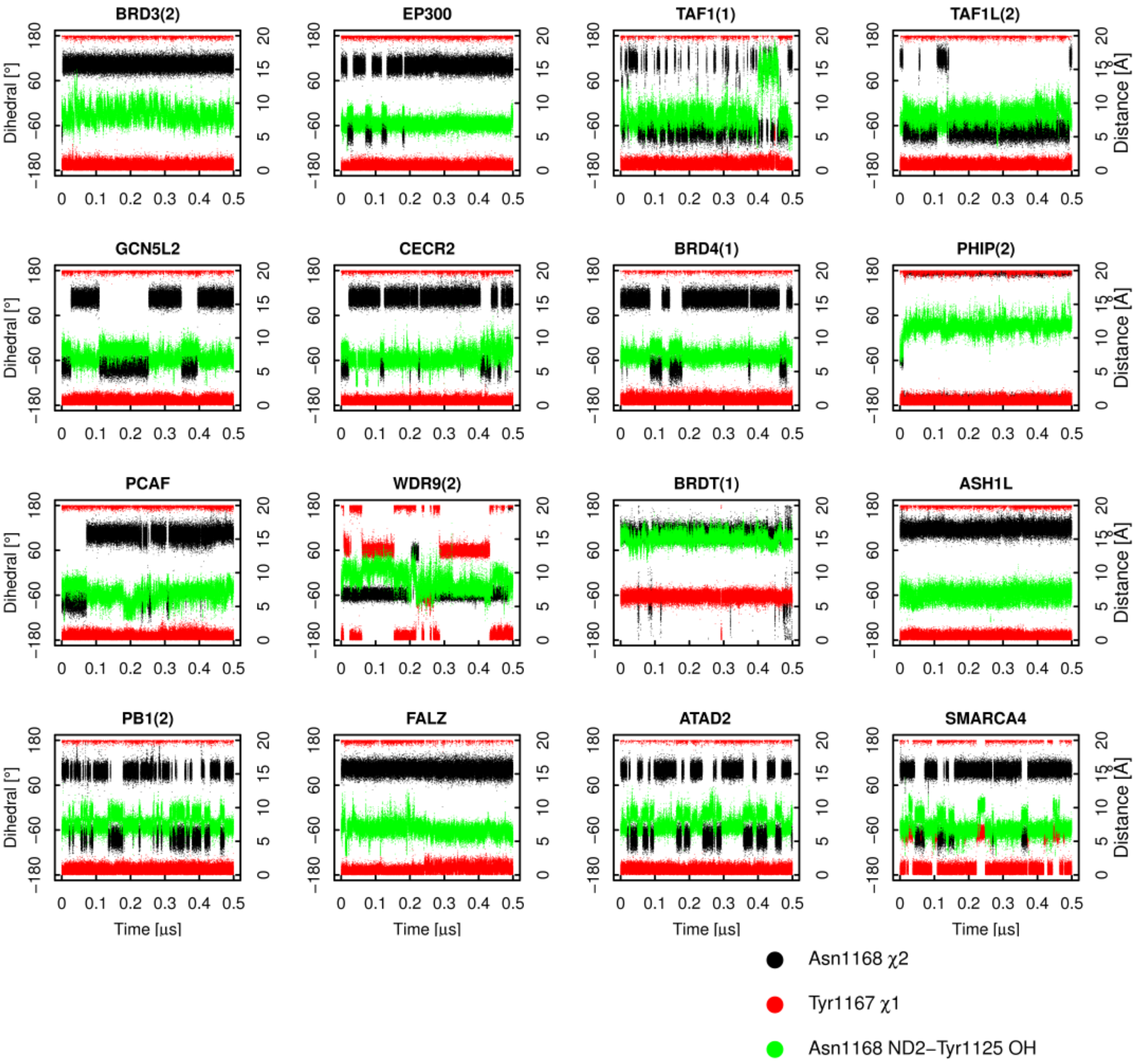


Figure S4: **Conserved Asn and Tyr/Phe: Time series of dihedrals and distances.** The χ_2 dihedral of the conserved Asn (i.e., Asn1168 of CREBBP) is shown (black). Note that for PHIP(2) and WDR9(2) χ_1 of Thr equivalent to Asn1168, and for ASH1L χ_2 of Tyr equivalent to Asn1168 are shown. The χ_1 of Tyr/Phe preceding in sequence the conserved Asn (i.e., Tyr1167 in CREBBP) is also shown (red). The distance between the side chain ND2 of the conserved Asn and the hydroxyl oxygen of the ZA-loop Tyr pointing inside the binding groove (i.e., Tyr1125 in CREBBP) is also shown (green); in case of a Thr or Tyr in position of the conserved Asn, Thr OG1, resp. Tyr OH were used instead of Asn ND2. Note that the ZA-loop Tyr is conserved in all the 20 bromodomains investigated here (see Figure S1).

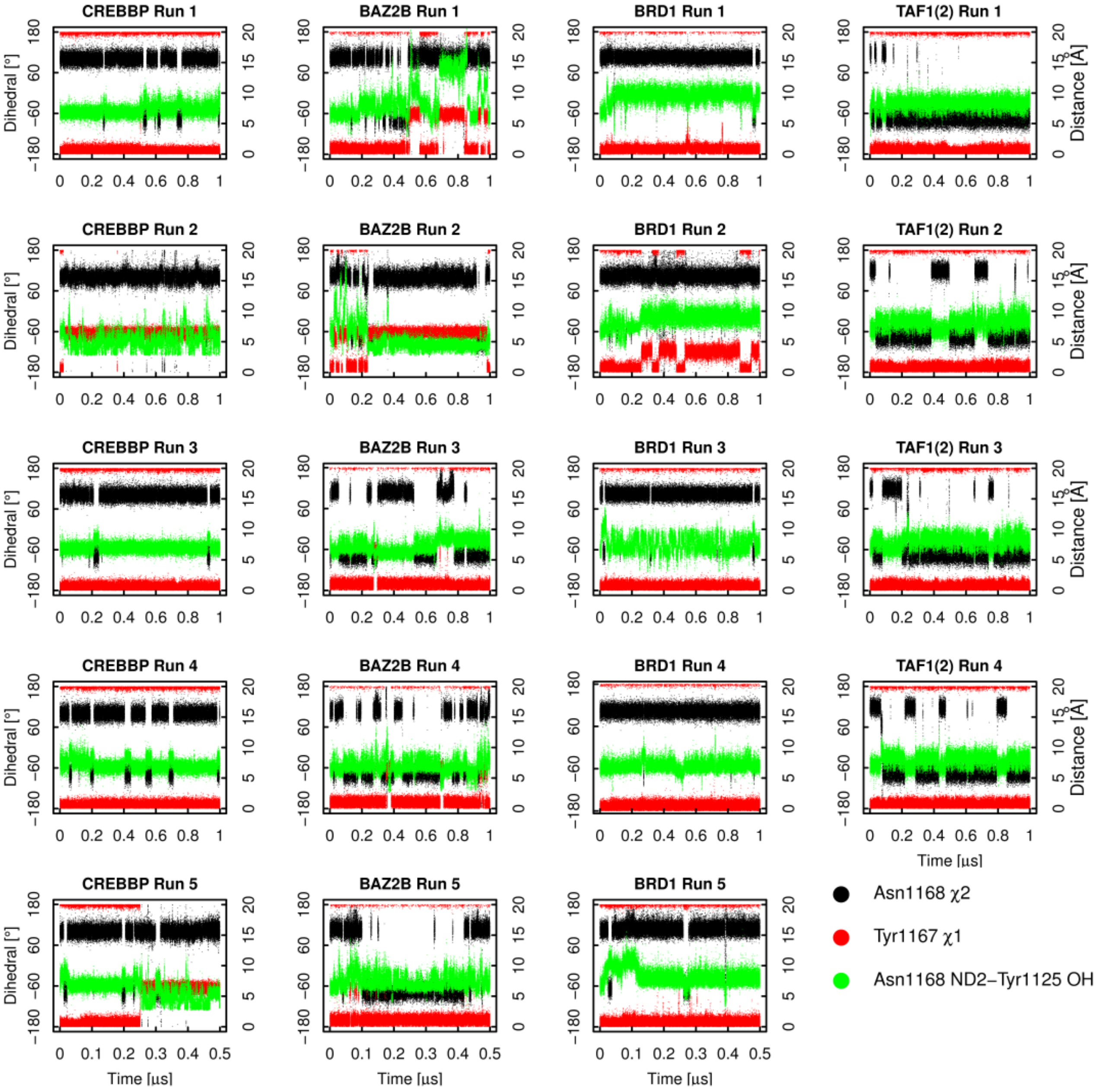


Figure S5: **Conserved Asn and Tyr/Phe: Time series of dihedrals and distances.** Same as Figure S4 for the four bromodomains for which multiple MD simulations were carried out.

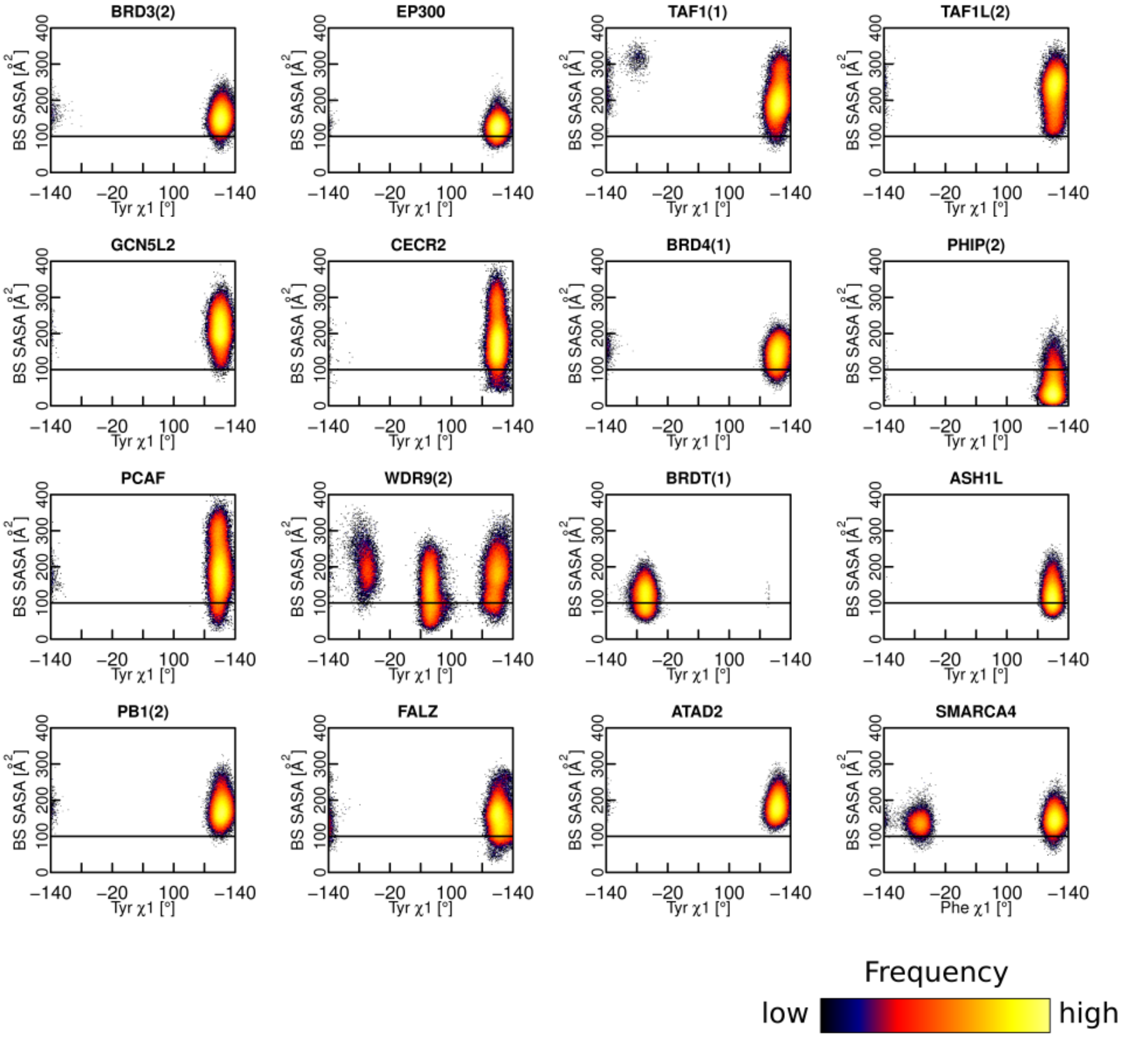


Figure S6: **Binding site accessible surface.** Two-dimensional histograms of binding site accessible molecular surface (SASA) versus χ_1 dihedral of Tyr/Phe preceding the conserved Asn in the BC-loop. Coloring according to frequencies follows a logarithmic scale.

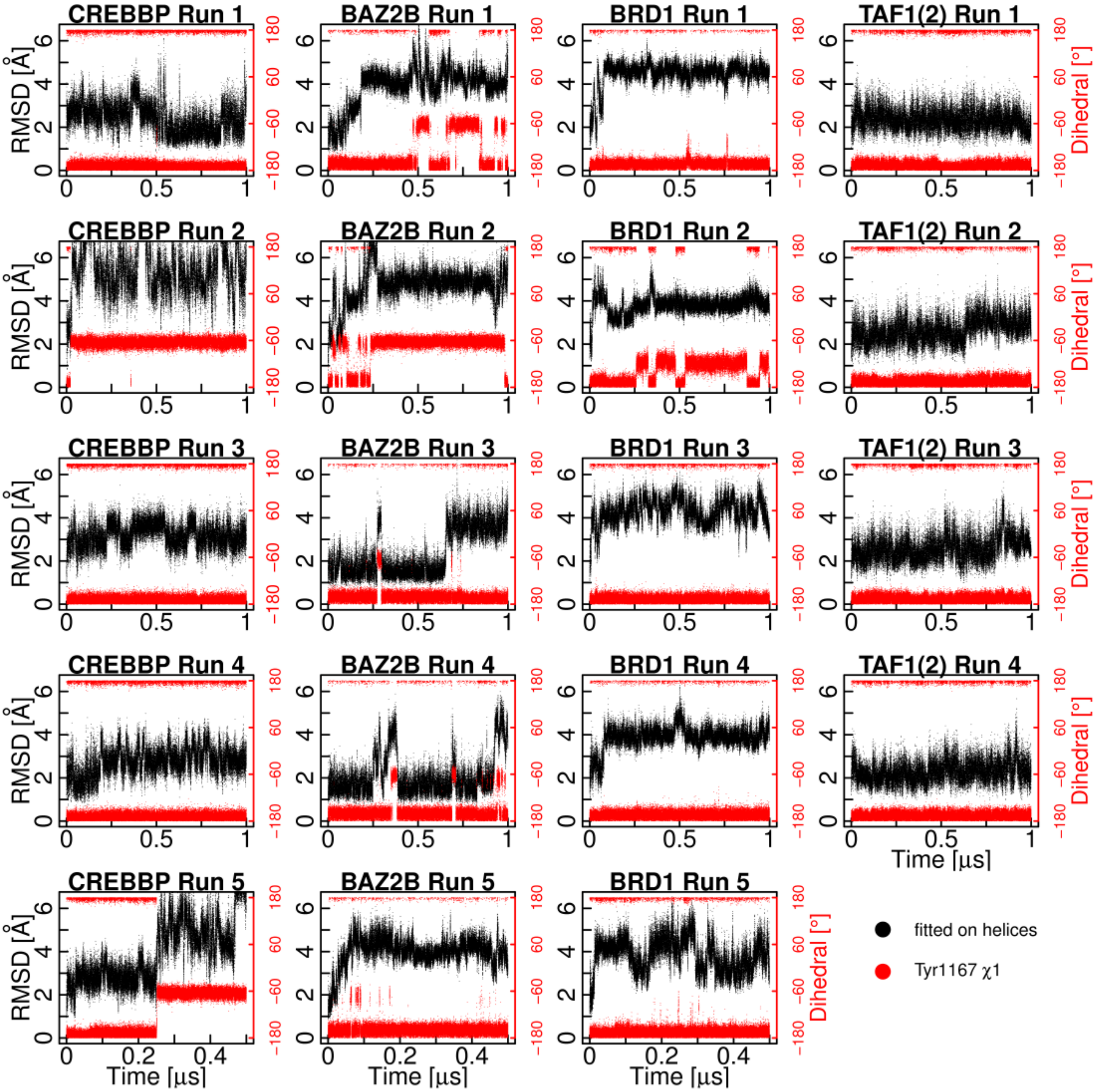


Figure S7: **Flexibility of ZA-loop and side chain rotation of the conserved Tyr/Phe in BC-loop.** Each panel corresponds to an independent simulation (of the bromodomain indicated in the top) and shows the time series of RMSD of the C_α atoms of the ZA-loop (black, y-axis on the left) and the χ_1 dihedral angle of the conserved Tyr/Phe in the BC-loop (red, y-axis on the right). The RMSD of the ZA-loop from the energy-minimized crystal structure was calculated upon structural overlap of the MD snapshots using the C_α atoms of the four α-helices.

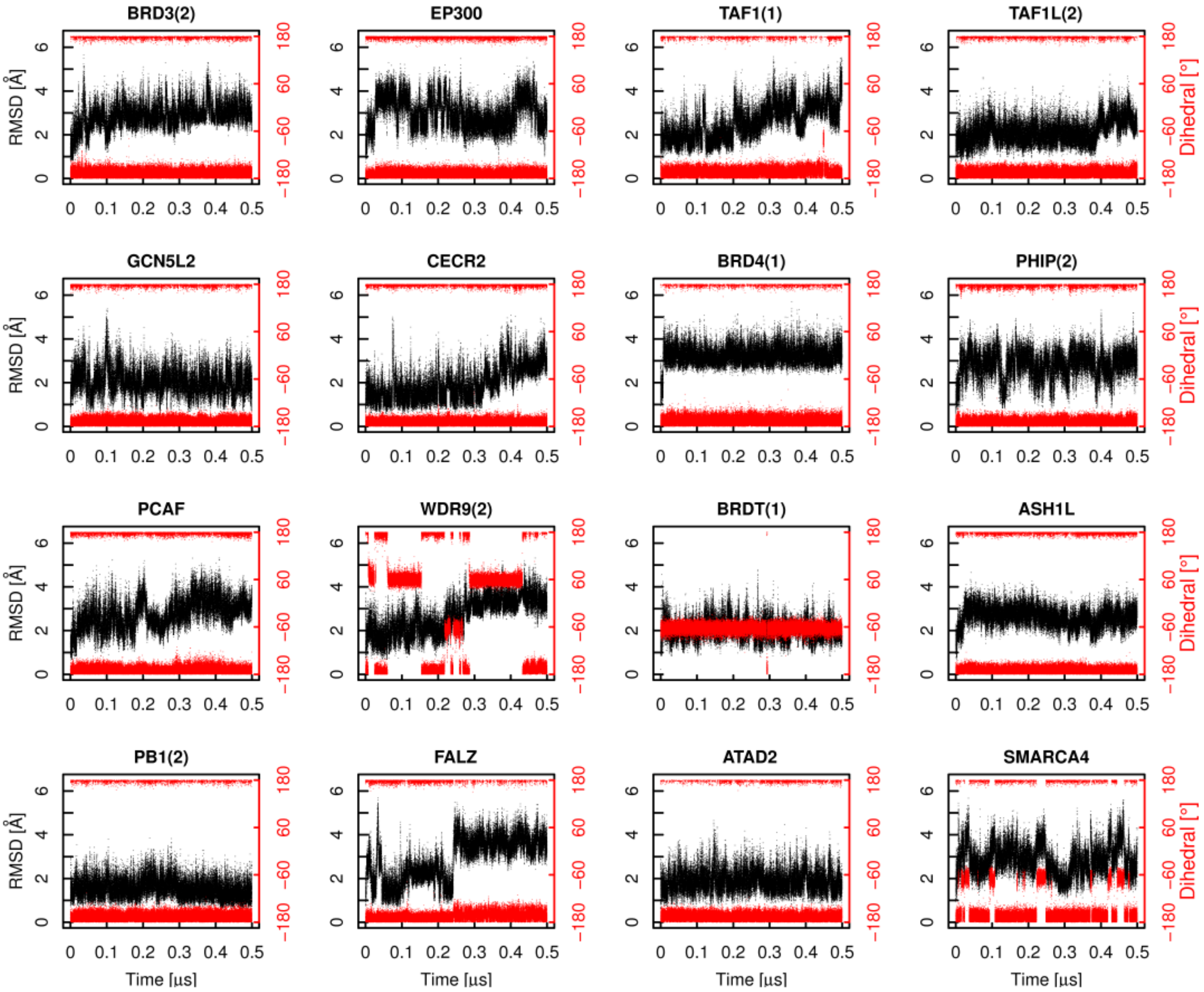


Figure S8: **Flexibility of ZA-loop and side chain rotation of the conserved Tyr/Phe in BC-loop.** See caption of Figure S7.

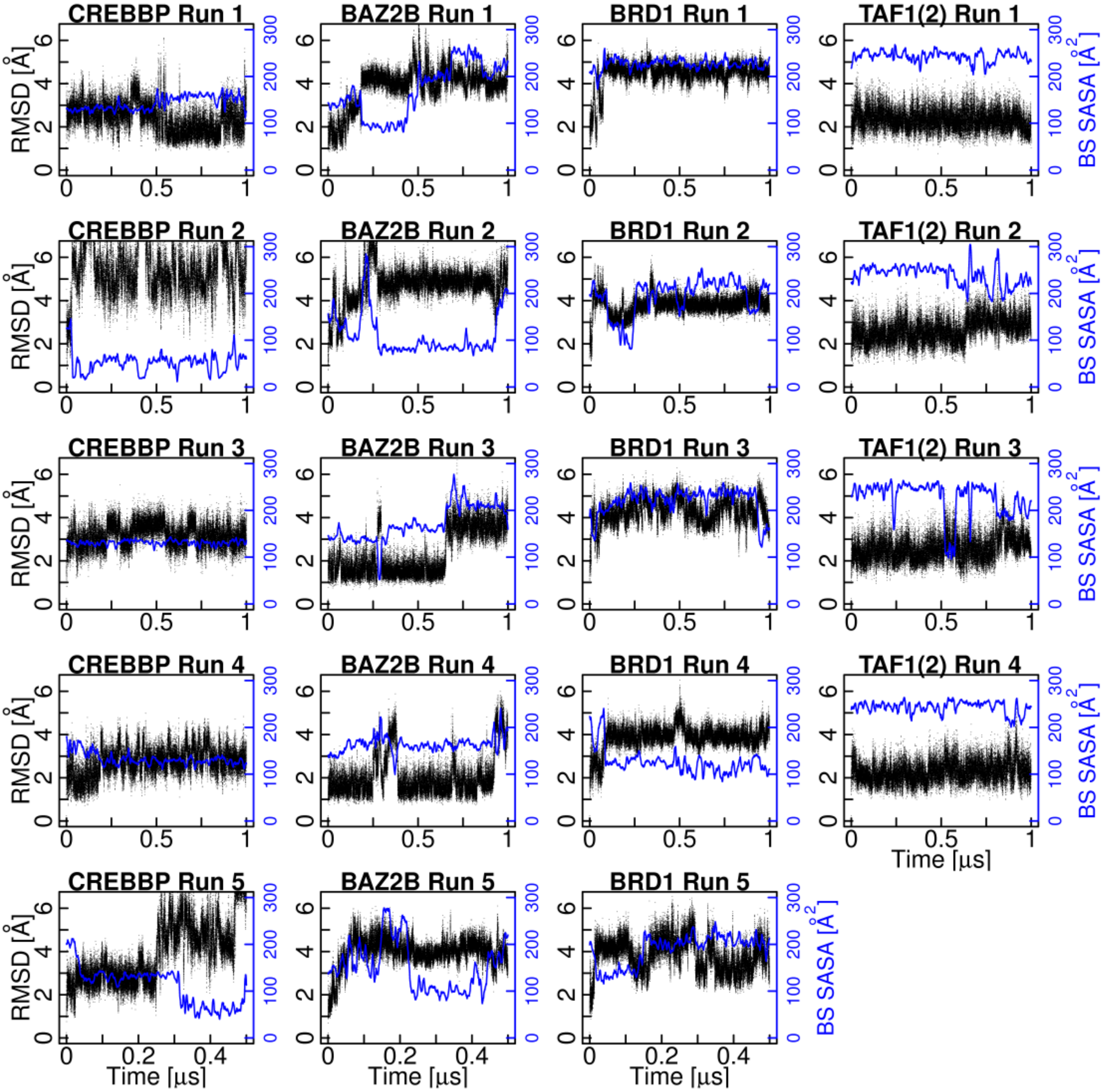


Figure S9: **Flexibility of ZA-loop and changes in binding site SASA.** Each panel corresponds to an independent simulation (of the bromodomain indicated in the top) and shows the time series of RMSD of the C_α atoms of the ZA-loop (black, y-axis on the left) and the binding site SASA (blue, y-axis on the right, running average calculated over time intervals of 10 ns for clarity). The RMSD of the ZA-loop from the energy-minimized crystal structure was calculated upon structural overlap of the MD snapshots using the C_α atoms of the four α-helices.

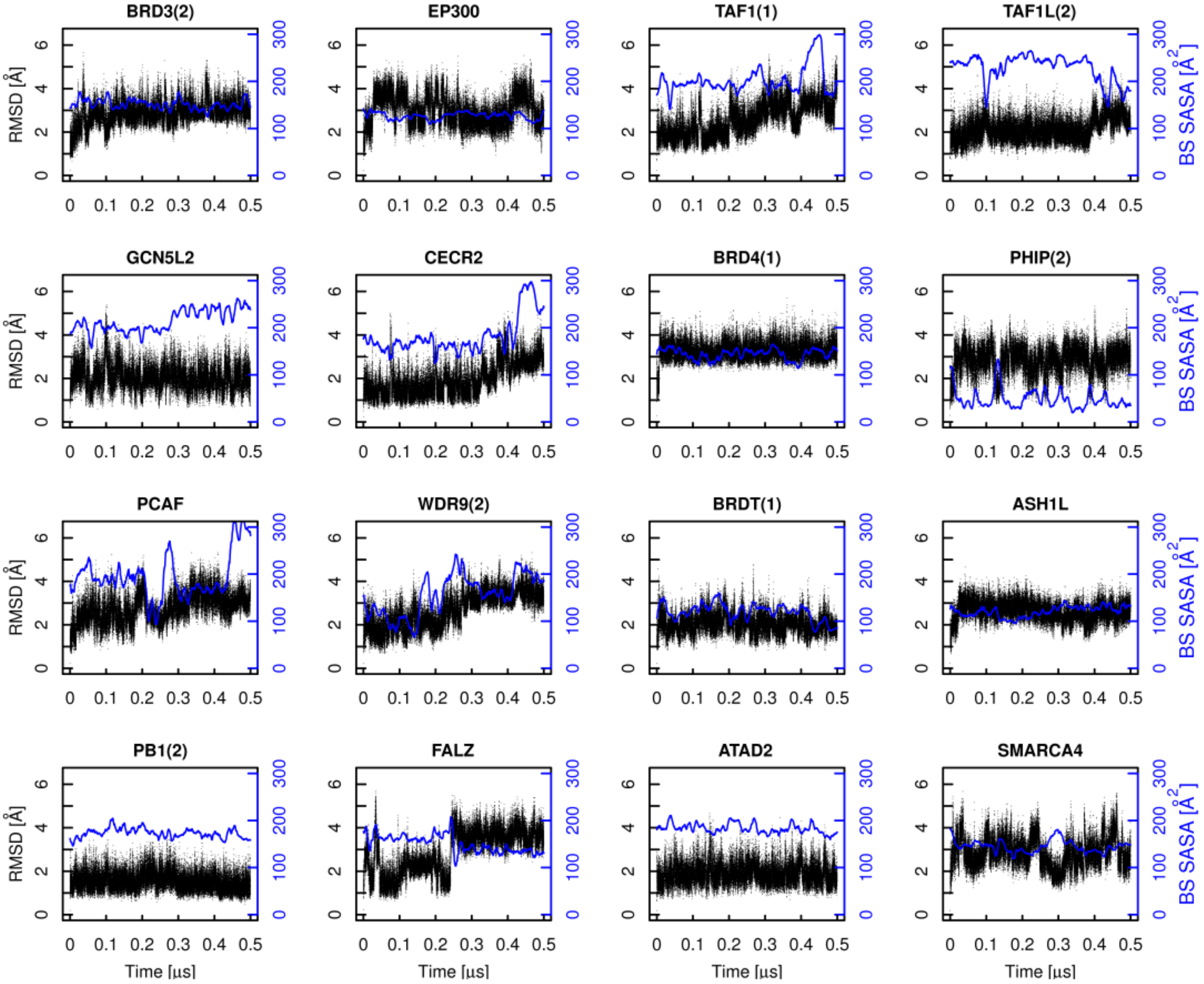


Figure S10: **Flexibility of ZA-loop and changes in binding site SASA.** See caption of Figure S9.

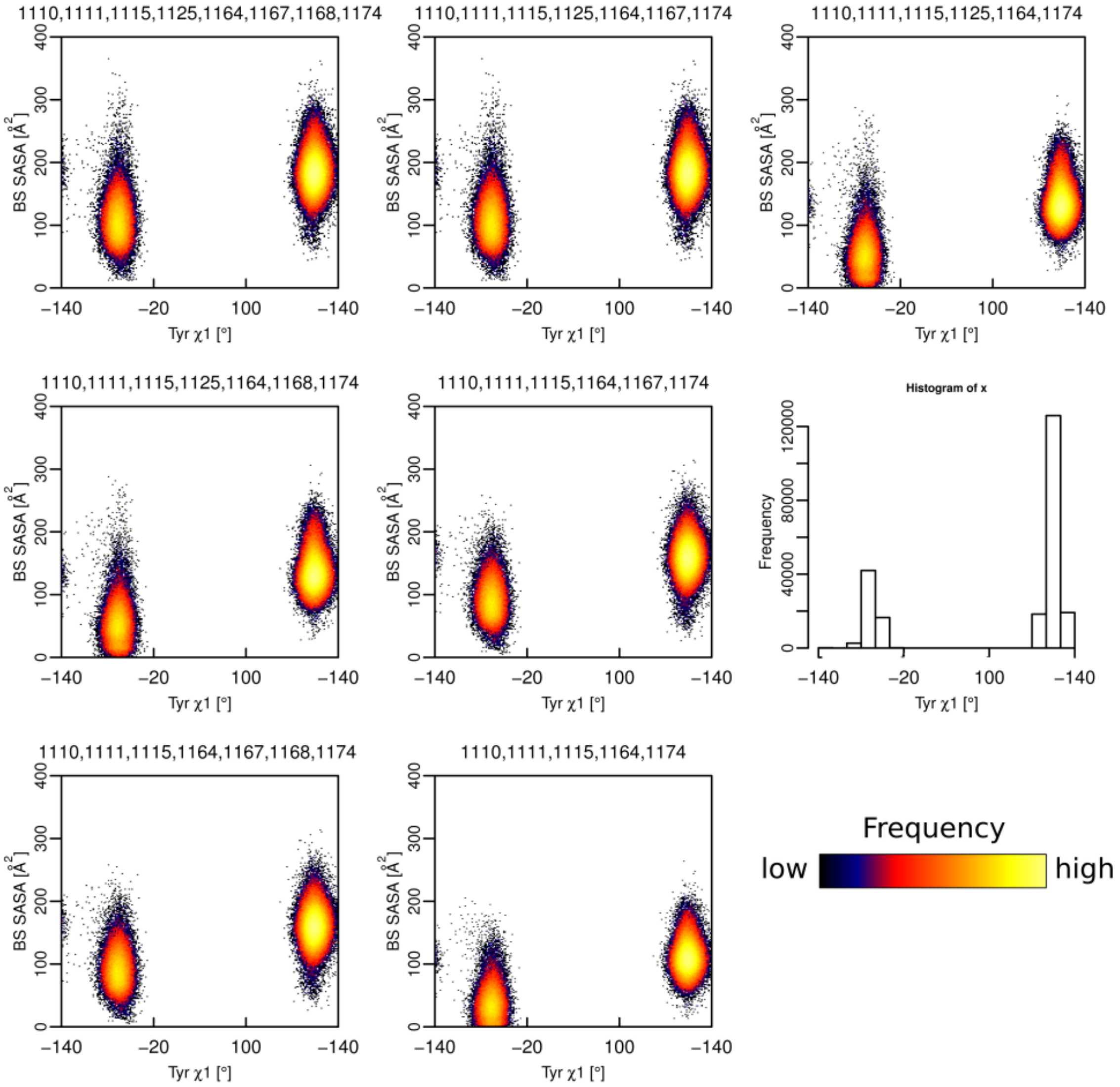


Figure S11: **Robustness of two-dimensional histograms.** The very similar two-dimensional histograms of binding site SASA vs. χ_1 of Tyr1167 for CREBBP, calculated using seven different selections of binding site atoms, demonstrate robustness. The subset of binding site residues used for each of the seven panels are the side chain heavy atoms of the residues whose numbers are mentionend above each panel. Coloring according to frequencies follows a logarithmic scale. The one-dimensional histogram of the χ_1 of CREBBP Tyr1167 is also shown (middle right).

References

[1] Filippakopoulos, P., Picaud, S., Mangos, M., Keates, T., Lambert, J.-P., Barsyte-Lovejoy, D., Felletar, I., Volkmer, R., Müller, S., Pawson, T., Gingras, A.-C., Arrowsmith, C. H., and Knapp, S. Histone recognition and large-scale structural analysis of the human bromodomain family. Cell 149(1):214–231, 2012.

[2] Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32(5):1792–1797, 2004.

[3] Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P., McWilliam, H., Valentin, F., Wallace, I., Wilm, A., Lopez, R., Thompson, J., Gibson, T., and Higgins, D. Clustal W and Clustal X version 2.0. Bioinformatics 23(21):2947–2948, 2007.

Chapter 6

Structured water molecules in bromodomains binding site can be displaced by cosolvent

Huang, D., Rossini, E., Steiner, S. and Caflisch, A.; Article in press at *ChemMedChem*

DOI: 10.1002/cmdc.201300156

Structured Water Molecules in the Binding Site of Bromodomains Can Be Displaced by Cosolvent

Danzhi Huang,* Emanuele Rossini, Sandra Steiner, and Amedeo Caflisch*[a]

Bromodomains are α-helical bundles of approximately 110 residues that recognize acetylated lysine side chains mainly on histone tails. Bromodomains are known to play an important role in cancer and inflammation, and as such, significant efforts are being made to identify small-molecule inhibitors of these epigenetic reader proteins. Here, explicit solvent molecular dynamics (MD) simulations of two bromodomains (BAZ2B and CREBBP) are used to analyze the water molecules that seem to be conserved at the bottom of the acetyl-lysine binding site in most crystal structures of bromodomains. The MD runs suggest that the occupancy of the structured water molecules is influenced by conformational transitions of the loop that connects helices Z and A. Additional simulations in the presence of 50 molecules of cosolvent (i.e., 440 mᴍ of dimethylsulfoxide, methanol, or ethanol) indicate that some of the structured water molecules can be displaced transiently. The residence time in the acetyl-lysine binding site is calculated to be about 1 ns, 2–5 ns, and 10–30 ns for methanol, ethanol, and dimethylsulfoxide, respectively, while the affinity of the three cosolvents for BAZ2B and CREBBP is in the range of 50–500 mᴍ. The results described have implications for ligand design, suggesting that only structured water molecules that do not exchange with cosolvent should be maintained in crystal structures used for docking campaigns, and that hydroxy substituents should be incorporated in the ligand so as to map the structured water molecules replaced by (m)ethanol.

Introduction

Bromodomains bind acetylated lysine side chains mainly in histones and are thus involved in the regulation of transcription.[1,2] As of today, 46 human proteins have been reported to include one or more of the 61 different bromodomains, with up to six bromodomains per protein.[3] As some bromodomains play an important role in cancer and inflammation,[4–6] there is an intense search for small-molecule inhibitors able to interfere in the process of reading acetylated lysine.[7] The first three-dimensional structure of a human bromodomain was solved by nuclear magnetic resonance spectroscopy.[8] Since then, the crystal and/or solution structures of more than 40 human bromodomains have been published.[9,10] All available structures show a conserved four-helix bundle topology (Figure 1 a) in which the ZA-loop and BC-loop connect the first two α helices (Z and A) and last two α helices (B and C), respectively.[10,11] The acetyl-lysine binding site is very similar in all structures of bromodomains.[10] It is substantially hydrophobic but contains at its bottom several well-ordered water molecules (Figure 1 b), which are present in most crystal structures of bromodomains.[7,10] In striking contrast to the abundance of available three-dimensional structures, to date, there have only been a few computational studies reported in the literature on bromodomains.[12–14] In particular, it seems that the dynamic properties of human apo bromodomains have not been investigated yet by atomistic simulations except for a very recent simulation study of 20 bromodomains.[15] On the other hand, several approaches to map (co)solvent to protein surfaces have been developed using protocols based on energy minimization[16–20] and more recently molecular dynamics (MD).[21–28]

Here, explicit solvent MD simulations are performed to analyze the occupancy of the conserved water molecules in the acetyl-lysine binding site and their potential displacement by dimethylsulfoxide (DMSO) or alcohol cosolvents ethanol and methanol. DMSO and ethanol differ mainly in the capability to act as a hydrogen-bond donor—only ethanol is able to do so. Methanol was selected for study because it differs in a single methylene group from ethanol, which provides information on size effects and hydrophobic-effect-dominated binding. Two bromodomains belonging to two different families are used in the simulations: the bromodomain adjacent to zinc finger domain 2B (BAZ2B) and the binding protein of the cAMP response element binding protein (CREBBP). There are several examples of explicit solvent MD simulations of small molecules binding to proteins.[22,23,29–33] On the other hand, the present work is the first simulation analysis of cosolvent binding to bromodomains. The simulation results reveal that the conformational changes of the ZA-loop are likely to have an influence on water occupancy. Moreover, the results suggest that some of the water molecules that are conserved in the available crystal structures of bromodomains can be transiently replaced by cosolvent molecules. This information is very useful

---

[a] *D. Huang, E. Rossini, S. Steiner, Prof. A. Caflisch*
Department of Biochemistry, University of Zürich
Winterthurerstrasse 190, 8057 Zürich (Switzerland)
E-mail: dhuang@bioc.uzh.ch
caflisch@bioc.uzh.ch

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cmdc.201300156.

WILEY ONLINE LIBRARY © 2013 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

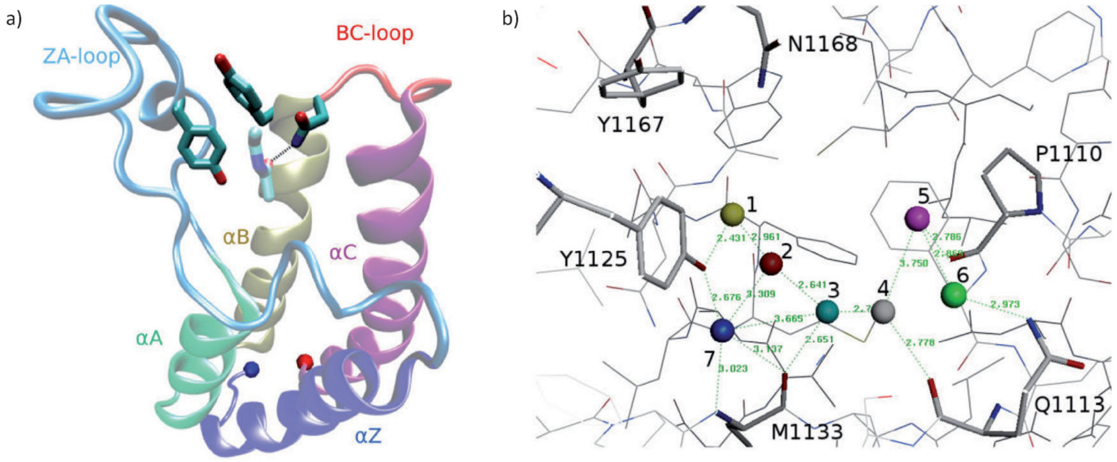


Figure 1. Bromodomain structure and structured water molecules in the bottom of the acetyl-lysine binding site. a) Ribbon illustration of the crystal structure of the complex between CREBBP and acetyl-lysine (PDB code: 3P1C[10]). Each of the four α-helices and two binding site loops is displayed with a different color. The side chains of the conserved residues Tyr 1125 of the ZA-loop, and Tyr 1167–Asn 1168 of the BC-loop are emphasized (sticks) together with the acetyl-lysine ligand (sticks, light colors). The N and C termini are shown with a blue and red sphere, respectively. The structure of BAZ2B (not shown) is very similar to that of CREBBP. The conserved residues in BAZ2B are Tyr 1901 in the ZA-loop, and Phe 1943–Asn 1944 in the BC-loop. b) Positions of structured water molecules. Water molecules 1–6 are the same as described in Ref. [7].

for hit identification by high-throughput docking and hit optimization based on docking models.[34–45] Our simulation results suggest to keep only the structured water molecules that do not exchange with cosolvent in high-throughput docking campaigns. Moreover, the position of structured water molecules that are replaced by (m)ethanol can be used to design hydroxy substituents of initial hits, which can improve the affinity by up to two orders of magnitude.[46–48]

Results

Simulations without cosolvent

For each of the structured water molecules, the occupancy was calculated using a threshold value for the distances between the water oxygen atom and three (mainly) backbone atoms in the acetyl-lysine binding site (Table S1 in the Supporting Information). The positions of water molecules 3 and 7 have the highest occupancy in the simulations of both bromodomains, while the position of water molecule 5 has the lowest occupancy (Figure 2). These simulation results are consistent with the degree of burial observed in the crystal structures,[7] which is maximal for water molecules 3 and 7. The time evolution of the water occupancy and root mean square deviation (RMSD) of the ZA-loop from its position in the crystal structure show a similar pattern (Figure 2), which indicates that the presence of the conserved water molecules is influenced by the motion of the ZA-loop. The structured water molecules connected by hydrogen bonds (i.e., water molecules 1–3) show similar temporal patterns of occupancy. In other words, these three water molecules are either all present or all absent along the simulation. The same behavior is observed for water molecules 5 and 6, which are also connected by a hydrogen bond. On the other hand, water molecule 4 can have temporal evolution of the occupancy similar as water molecules 1–3 or 5–6.

Simulations with cosolvent

These simulations were started from random positions of 50 cosolvent molecules in the bulk water and without any cosolvent molecule in the acetyl-lysine binding site. The maps of cosolvent occupancy (Figure 3) indicate that DMSO, ethanol, and methanol bind preferentially to the acetyl-lysine binding site. From these maps, it also emerges that methanol can bind to most of the bromodomain surface. Methanol is less specific than ethanol and DMSO because it is the smallest of these three cosolvents, and furthermore, in contrast to DMSO, it can also act as a hydrogen-bond donor.

The time series of the distance of the DMSO molecule closest to the acetyl-lysine binding site of CREBBP demonstrates that the MD sampling is adequate as multiple binding and unbinding events are observed and by different DMSO molecules in the 0.5 μs time scale of the MD runs (Figure 4). Cluster analysis of the MD trajectories show that the most populated binding modes of DMSO and (m)ethanol are consistent with the crystal structure of DMSO bound to CREBBP (PDB code: 3P1E[10]) and 1,2-ethanediol bound to the second bromodomain of TAF1 (PDB code: 3UV4[10]), respectively (Figure 4). (Note that there are no crystal structures of CREBBP and BAZ2B with alcohol cosolvent in the acetyl-lysine binding pocket and no crystal structure of DMSO–BAZ2B.) In their most populated binding mode, DMSO, ethanol, and methanol accept a hydrogen bond from the side chain of the same asparagine residue that is involved in binding the natural ligand acetyl-lysine (Figure 4).

© 2013 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

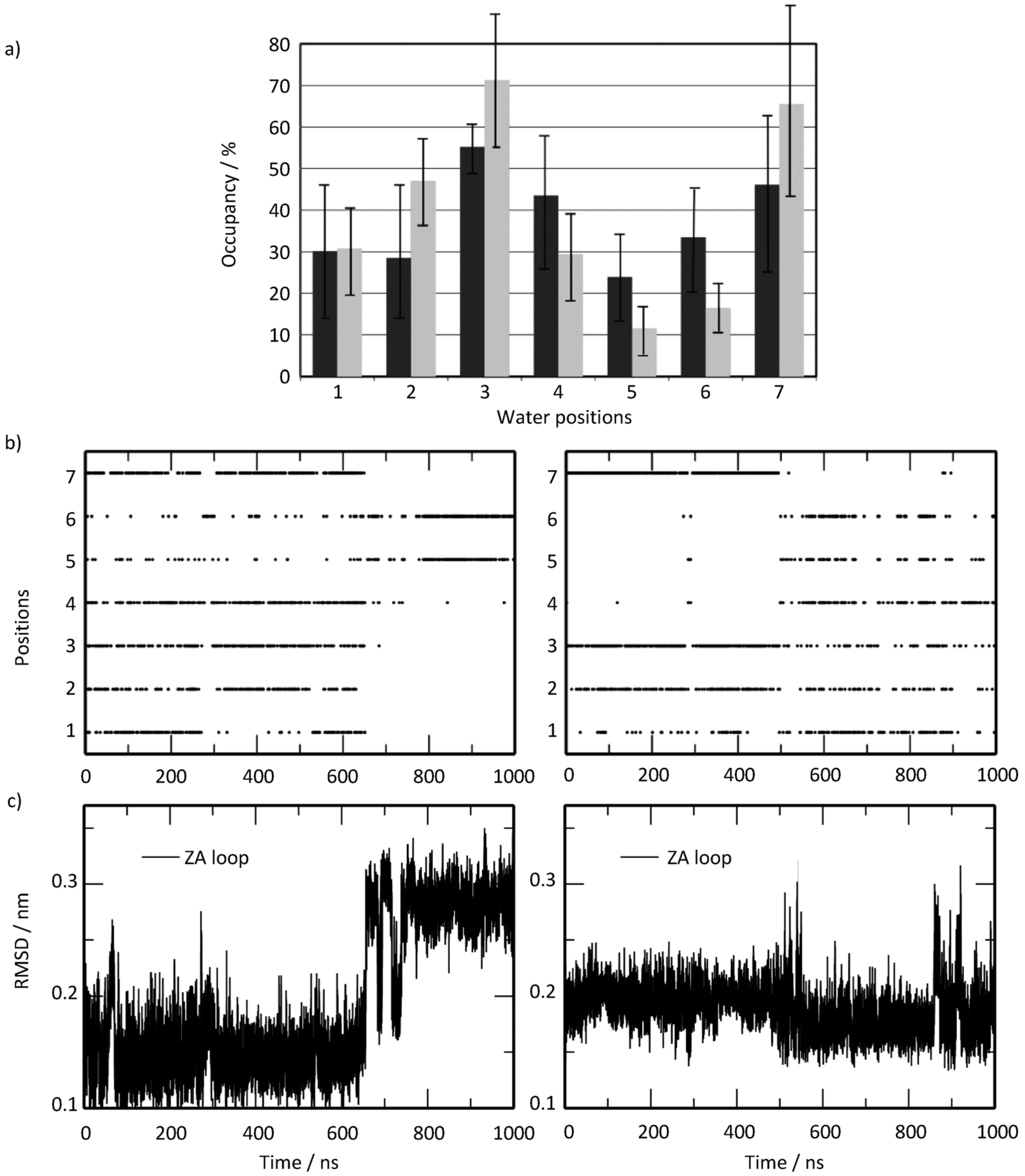


Figure 2. Water occupancy in the simulations without cosolvent. a) Percentage of water occupancy at each of the seven positions shown in Figure 1 b. Each bar shows the average and standard deviation calculated over four independent simulations without cosolvent for a total of 4 μs for each of the two proteins (■ BAZ2B; ■ CREBBP). b) Time series of occupancy of positions of structured water molecules along one of the four 1 μs MD runs for BAZ2B (left) and CREBBP (right). c) Time series of RMSD from the crystal structure of the carbon atoms in the ZA loop upon structural overlap of the carbon atoms of residues in nonflexible parts of the bromodomain. These two panels show results for the same MD runs shown in panel b.

Discussion

As mentioned above, the analysis of the occupancy of the structured water molecules in the presence of cosolvent is useful for docking and hit optimization. The seven structured water molecules are present in the most populated binding mode of the cosolvent while they can be displaced for more buried binding modes. The time series of water or cosolvent occupancy (Figure 5; see also Figure S1 in the Supporting Information) show that the cosolvent displaces mainly the structured water molecules at positions 1 and 5. In addition, there is transient displacement of water molecules 2–4 and 6 by methanol and ethanol. In contrast, the deeply buried water molecule 7 is displaced much less frequently by cosolvent. For example, with BAZ2B, the displacement of individual structured water molecules by ethanol is observed in 10%, 0.3%, 0.5%, 0.1%, 4%, 0.3%, and 0% of the MD sampling for positions 1–7, respectively (for other values of cosolvent occupancy, see Table S2 in the Supporting Information).

The sampling of multiple binding and unbinding events allows one to extract the kinetics directly from the MD simulations (Table 1). The binding and unbinding times are calculated from the fitting of the respective cumulative distribution (Figure S2 in the Supporting Information) as in a previous MD study of the (un)binding of cosolvent molecules from the FK506 binding protein.[32,33] The binding times are similar

© 2013 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

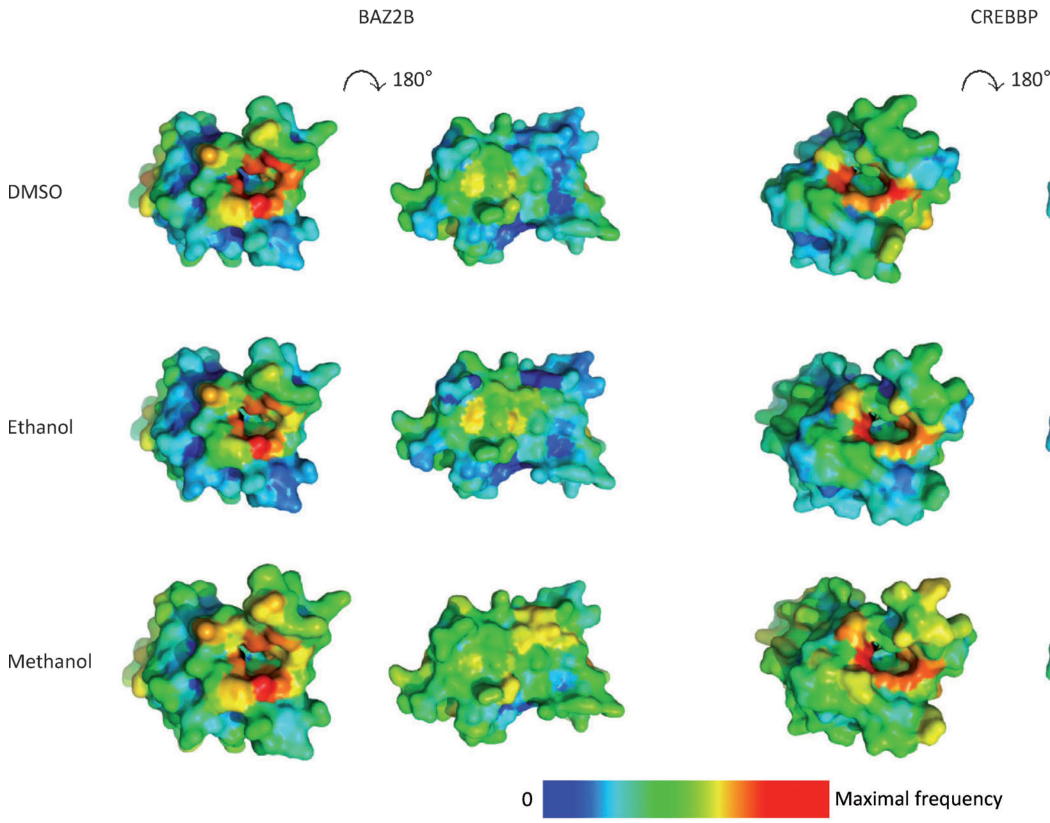


Figure 3. Maps of cosolvent/bromodomain contact frequencies. The surface of the protein is colored according to the percentage of MD snapshots in which the center of mass of any of the 50 molecules of cosolvent is within 0.7 nm of any nonhydrogen atom of the protein. The structures in the first and third columns are oriented such that the acetyl-lysine binding pocket is perpendicular to the plane of the image. The structures in the second and fourth columns are obtained by 180° rotation around a horizontal axis in the plane of the image. These maps show that, for all three cosolvents and both bromodomains, the acetyl-lysine binding pocket is the main binding site.

(factor of five between the slowest and fastest) for the three cosolvents and two bromodomains, which is a consequence of diffusion-limited binding as the three cosolvents used in this study have similar size and polarity. Methanol, the smallest of the three cosolvents investigated here, has the lowest occupancy of the acetyl-lysine binding site for both bromodomains, which is due to its short residence times (about 1 ns). DMSO and ethanol have higher occupancy than methanol because of longer residence time (2.4–26 ns) and similar binding time. The slower unbinding of DMSO and ethanol with respect to methanol might be a consequence of their slightly larger size and/or hydrophobicity. Each of the three cosolvents has similar unbinding times for the two bromodomains. The only exception is the three times longer unbinding time of DMSO from CREBBP compared with BAZ2B. It is interesting to note that this simulation result is consistent with the higher affinity of DMSO for CREBBP than for BAZ2B (Table 1), as measured by a peptide-displacement assay[4,49] with AlphaScreen technology.[50] The experimental results indicate how strongly DMSO competes with an acetylated histone peptide that is a known ligand of the corresponding bromodomain. Note that the competition experiments do not measure directly the cosolvent dissociation constant and therefore a perfect agreement between experimental measurements and dissociation constants derived from the simulations is not expected.

Conclusions

The simulations of bromodomains in pure water show that the occupancy of the conserved water molecules depends on the orientation of the ZA-loop. In the simulations with cosolvent, most of the structured water molecules at the bottom of the acetyl-lysine binding pocket can be transiently displaced by ethanol or methanol. The heterogeneous occupancy of the structured water molecules revealed by the MD simulations is helpful for ligand design, for example, in deciding which of the structured water molecules to take into account for high-throughput docking campaigns. Furthermore, the water molecules displaced by (m)ethanol can be used directly in ligand design, where the inclusion of hydroxy substituents in the ligand so as to map the positions of these displaced structured water molecules could give rise to an increase in binding affinity.

Experimental Section

Simulation protocols: The coordinates of CREBBP and BAZ2B were downloaded from the protein database (PDB codes: 3DWY[10] and 3G0L[2], respectively).[51] To reproduce neutral pH conditions, the side chains of aspartates and glutamates were negatively charged, those of lysine and arginine residues were positively charged, and

© 2013 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

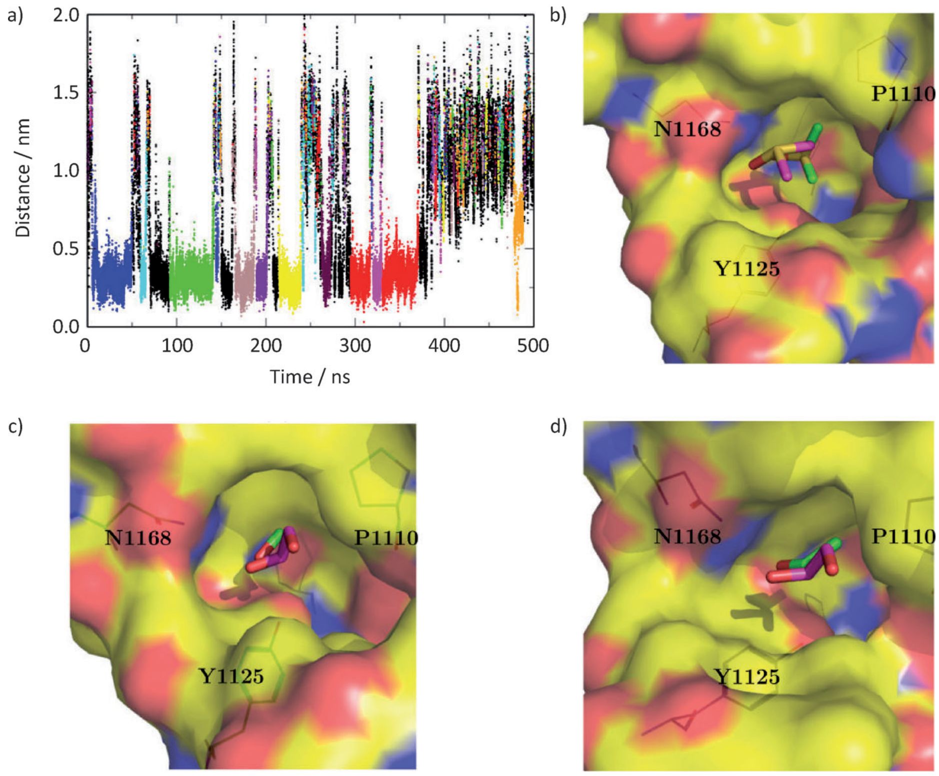


Figure 4. Reversible binding and major binding modes. a) Time series of distance between the center of the acetyl-lysine binding site in CREBBP and the closest DMSO molecule. Different DMSO molecules are "labeled" by different colors. b) The most populated binding mode of DMSO from the MD simulations of CREBBP. DMSO is shown by sticks colored by atomic element (O: red; S: yellow; C: green (pose obtained by MD) or magenta (binding mode in crystal structure PDB code: 3P1E[10]). The surface of CREBBP is colored by atomic element (C: yellow; N: blue; O: red). The same is shown for c) methanol and d) ethanol. The pose obtained by MD for methanol and ethanol is similar to the binding mode of 1,2-ethanediol (C atoms in magenta) in the second bromodomain of TAF1 (PDB code: 3UV4[10]).

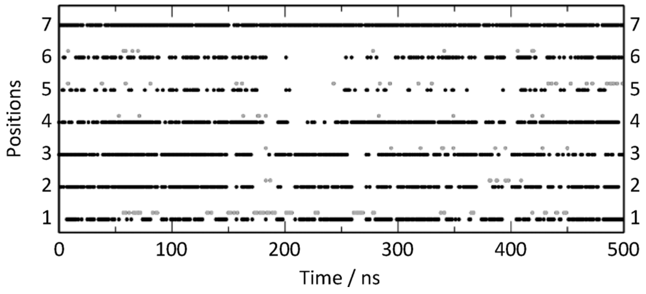


Figure 5. Water displacement by cosolvent: Time series of occupancy of ethanol (grey dots) or water (black dots) at each of the seven positions of structured water molecules along one of the two simulations of BAZ2B carried out in the presence of 50 molecules of ethanol molecules.

the histidine side chains were neutral. For each bromodomain, the crystal water molecules close to the binding site and with low B factor values were kept while the remaining water molecules were deleted. Subsequently, the structure was solvated in a water box, the size of which was chosen to have a minimal distance of 1.1 nm between the boundary and any atom of the protein. Fifty molecules of methanol, ethanol, or dimethylsulfoxide (DMSO) corresponding to a concentration of 440 mM were placed randomly in the bulk water. Solvation water molecules within 0.24 nm of any heavy atom of the protein or the cosolvent were removed. Sodium and chloride ions, the parameters for which are from the CHARMM PARAM22 force field,[52] were added randomly by replacing noncrystallographic water molecules to compensate for the total charge of each bromodomain and to approximate an ionic strength of about 150 mM. The MD simulations were carried out with Gromacs 4.5.5[53] using the CHARMM PARAM22 force field[52] and the TIP3P model of water.[54] Parameters for the three cosolvents were determined according to the CHARMM general force field.[55]

Periodic boundary conditions were applied and electrostatic interactions were evaluated using the particle-mesh Ewald summation method.[56] The van der Waals interactions were truncated at a cutoff of 1.0 nm. The MD simulations were performed at constant temperature (310 K) using the velocity rescaling thermostat and constant pressure (1 atm).[57] The LINCS algorithm was used to fix the covalent bonds involving hydrogen atoms, and a time step of 2 fs was used in all runs. Before each MD simulation, the whole system was relaxed by energy minimization until the maximal force on individual atoms was smaller than 100 kJ mol^{-1} nm^{-1}. After the minimization, initial atomic velocities, corresponding to the Boltzmann distribution at a temperature of 310 K, were randomly assigned and a constrained MD run was performed for 1 ns. During the constrained simulations, heavy atoms of the protein were fixed to their initial positions with a force constant of 1000 kJ mol^{-1} nm^{-1}. The constraints were released, and the system was equilibrated for 10 ns before data collection for analysis. For each of the two bromodomains (CREBBP and BAZ2B), four independent 1 μs MD runs without cosolvent were carried out. In addition, two independent MD runs of 0.5 μs each were performed for each of the six cosolvent/bromodomain systems. Details of the simulation protocol are given in the Supporting Information.

Analysis: The analysis of the MD trajectories was carried out with CHARMM[58] and the MD analysis tool WORDOM.[59] Occupancy of each of the seven positions of the structured water molecules was determined using the triplets of binding site atoms listed in Table S1 in the Supporting Information.

Clustering according to pairwise RMSD of the cosolvent atoms was carried out using the leader algorithm as implemented in WORDOM. First, the structural overlap was determined using the bromodomain carbon atoms excluding those in the flexible regions, i.e., the ZA and BC loops and the five terminal residues. Upon overlap of the bromodomain structures, the RMSD was evaluated for the nonhydrogen atoms of the cosolvent molecule closest to the acetyl-lysine binding site. A threshold value of 0.2 nm for

© 2013 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 1. Thermodynamics and kinetics of cosolvent binding.

		Simulation results					Experimental data
Protein	Cosolvent	τ_{on} [ns]	τ_{off} [ns]	k_{off}/k_{on}[a] [mM]	Occupancy[b] [%]	K_D[c] [mM]	Affinity[d] [mM]
BAZ2B	methanol	1.1	0.9	543	54	375	–
	ethanol	0.7	2.4	124	82	97	–
	DMSO	2.6	8.9	128	67	218	347
CREBBP	methanol	0.5	1.3	176	55	360	–
	ethanol	1.0	4.5	94	86	71	–
	DMSO	2.1	26	36	75	147	18

[a] Dissociation constant evaluated as the quotient of k_{off} and k_{on} which are derived from fitting of the cumulative distributions of unbinding times and binding times, respectively (Figure S2 in the Supporting Information). The characteristic time of the slow phase of the double-exponential fitting is used to calculate the binding rate $k_{on} = 1/(\tau_{on}[\text{cosolvent}])$ and unbinding rate $k_{off} = 1/\tau_{off}$. The concentration of the cosolvent in the simulation box is 440 mM. The distance between the binding site center and the mass center of the cosolvent is used to monitor binding. The binding site center of CREBBP is defined by the geometric center of the carbon α atoms of residues Pro 1110, Pro 1114, Tyr 1125, and Asn 1168. The binding site center of BAZ2B is defined by the geometric center of the carbon α atoms of residues Pro 1888, Pro 1892, Tyr 1901 and Asn 1944. The threshold distances used for the calculations of binding and unbinding times are 0.4 nm and 0.8 nm, respectively. Results with different cutoff combinations are listed in Table S3 in the Supporting Information. Using binding cutoffs in the range 0.3–0.5 nm and unbinding cutoffs in the range 0.7–0.9 nm results in differences in K_D smaller than a factor of three. [b] The occupancy of the active site was determined using the same threshold for the distance between the binding site center and the mass center of the cosolvent as for the unbinding kinetics (0.8 nm). [c] Dissociation constant (K_D) calculated from occupancy of acetyl-lysine binding site using $K_D = [\text{cosolvent}]$ [unbound protein]/[bound protein] = [cosolvent] (100−occupancy)/occupancy. The robustness of the occupancy upon changes in the threshold is shown in Figure S3 in the Supporting Information. Using threshold values ranging from 0.5 to 0.9 nm results in differences in K_D smaller than a factor of two. [d] Experimental value of binding affinity measured by AlphaScreen peptide displacement assay.[4,49]

the RMSD from the cluster representative was used for assigning a snapshot to a cluster.

Acknowledgements

The authors thank Dr. Andrea Magno (University of Zürich) for interesting discussions. The simulations were carried out on the Schrödinger compute cluster at the University of Zurich (Switzerland). This work was supported financially by a grant from the Swiss National Science Foundation to A.C.

Keywords: epigenetic readers · fragment-based design · molecular dynamics · residence times · small-molecule binding

[1] L. Zeng, M.-M. Zhou, *FEBS Lett.* **2002**, *513*, 124–128.
[2] P. Filippakopoulos, S. Knapp, *FEBS Lett.* **2012**, *586*, 2692–2704.
[3] C. H. Arrowsmith, C. Bountra, P. V. Fish, K. Lee, M. Schapira, *Nat. Rev. Drug Discovery* **2012**, *11*, 384–400.
[4] P. Filippakopoulos, J. Qi, S. Picaud, Y. Shen, W. B. Smith, O. Fedorov, E. M. Morse, T. Keates, T. T. Hickman, I. Felletar, M. Philpott, S. Munro, M. R. McKeown, Y. Wang, A. L. Christie, N. West, M. J. Cameron, B. Schwartz, T. D. Heightman, N. L. Thangue, C. A. French, O. Wiest, A. L. Kung, S. Knapp, J. E. Bradner, *Nature* **2010**, *468*, 1067.
[5] J. Zuber, J. Shi, E. Wang, A. R. Rappaport, H. Herrmann, E. A. Sison, D. Magoon, J. Qi, K. Blatt, M. Wunderlich, M. J. Taylor, C. Johns, A. Chicas, J. C. Mulloy, S. C. Kogan, P. Brown, P. Valent, J. E. Bradner, S. W. Lowe, C. R. Vakoc, *Nature* **2011**, *478*, 524–528.
[6] M. A. Dawson, T. Kouzarides, B. J. Huntly, *N. Engl. J. Med.* **2012**, *367*, 647–657.
[7] D. S. Hewings, T. P. C. Rooney, L. E. Jennings, D. A. Hay, C. J. Schofield, P. E. Brennan, S. Knapp, S. J. Conway, *J. Med. Chem.* **2012**, *55*, 9393–9413.
[8] C. Dhalluin, J. E. Carlson, L. Zeng, C. He, A. K. Aggarwal, M.-M. Zhou, *Nature* **1999**, *399*, 491–496.
[9] S. Mujtaba, L. Zeng, M. Zhou, *Oncogene* **2007**, *26*, 5521–5527.
[10] P. Filippakopoulos, S. Picaud, M. Mangos, T. Keates, J.-P. Lambert, D. Barsyte-Lovejoy, I. Felletar, R. Volkmer, S. Müller, T. Pawson, A.-C. Gingras, C. H. Arrowsmith, S. Knapp, *Cell* **2012**, *149*, 214–231.
[11] L. R. Vidler, N. Brown, S. Knapp, S. Hoelder, *J. Med. Chem.* **2012**, *55*, 7346–7359.
[12] S. Pantano, A. Marcello, A. Ferrari, D. Gaudiosi, A. Sabò, V. Pellegrini, F. Beltram, M. Giacca, P. Carloni, *Proteins* **2006**, *62*, 1062–1073.
[13] F. Pizzitutti, A. Giansanti, P. Ballario, P. Ornaghi, P. Torreri, G. Ciccotti, P. Filetici, *J. Mol. Recognit.* **2006**, *19*, 1–9.
[14] K. D. Eichenbaum, Y. Rodriguez, M. Mezei, R. Osman, *Proteins Struct. Funct. Bioinf.* **2010**, *78*, 447–456.
[15] S. Steiner, A. Magno, D. Huang, A. Caflisch, *FEBS Lett.* **2013**, *in press*; DOI: 10.1016/j.febslet.2013.05.032
[16] A. Miranker, M. Karplus, *Proteins Struct. Funct. Genet.* **1991**, *11*, 29–34.
[17] A. Caflisch, A. Miranker, M. Karplus, *J. Med. Chem.* **1993**, *36*, 2142–2167.
[18] D. Joseph-McCarthy, J. Hogle, M. Karplus, *Proteins Struct. Funct. Bioinf.* **1997**, *29*, 32–58.
[19] C. Stultz, M. Karplus, *Proteins Struct. Funct. Genet.* **1999**, *37*, 512–529.
[20] M. Scarsi, N. Majeux, A. Caflisch, *Proteins Struct. Funct. Bioinf.* **1999**, *37*, 565–575.
[21] R. Baron, J. A. McCammon, *Biochemistry* **2007**, *46*, 10629–10642.
[22] J. Seco, F. J. Luque, X. Barril, *J. Med. Chem.* **2009**, *52*, 2363–2371.
[23] O. Guvench, A. D. MacKerell, Jr., *PLoS Comput. Biol.* **2009**, *5*, e1000435.
[24] K. W. Lexa, H. A. Carlson, *Q. Rev. Biophys.* **2012**, *45*, 301–343.
[25] R. Baron, J. A. McCammon, *Annu. Rev. Phys. Chem.* **2013**, *64*, 151–175.
[26] P. Setny, R. Baron, P. M. Kekenes-Huskey, J. A. McCammon, J. Dzubiella, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 1197–1202.
[27] K. W. Lexa, H. A. Carlson, *J. Chem. Inf. Model.* **2013**, *53*, 391–402.
[28] K. W. Lexa, H. A. Carlson, *J. Am. Chem. Soc.* **2011**, *133*, 200–202.
[29] D. Ekonomiuk, X.-C. Su, C. Bodenreider, S. P. Lim, G. Otting, D. Huang, A. Caflisch, *J. Med. Chem.* **2009**, *52*, 4860–4868.
[30] N. Basse, J. L. Kaar, G. Settanni, A. C. Joerger, T. J. Rutherford, A. R. Fersht, *Chem. Biol.* **2010**, *17*, 46–56.
[31] I. Buch, T. Giorgino, G. D. Fabritiis, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 10184–10189.
[32] D. Z. Huang, A. Caflisch, *ChemMedChem* **2011**, *6*, 1578–1580.
[33] D. Z. Huang, A. Caflisch, *PLoS Comput. Biol.* **2011**, *7*, e1002002.
[34] C. Poornima, P. Dean, *J. Comput.-Aided Mol. Des.* **1995**, *9*, 521–531.
[35] W. E. Minke, D. J. Diller, W. G. Hol, C. L. Verlinde, *J. Med. Chem.* **1999**, *42*, 1778–1788.
[36] R. L. Mancera, *J. Comput.-Aided Mol. Des.* **2002**, *16*, 479–499.
[37] C. de Graaf, P. Pospisil, W. Pos, G. Folkers, N. P. Vermeulen, *J. Med. Chem.* **2005**, *48*, 2308–2318.
[38] M. L. Verdonk, G. Chessari, J. C. Cole, M. J. Hartshorn, C. W. Murray, J. W. M. Nissink, R. D. Taylor, R. Taylor, *J. Med. Chem.* **2005**, *48*, 6504–6515.

© 2013 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

[39] C. Barillari, J. Taylor, R. Viner, J. W. Essex, *J. Am. Chem. Soc.* **2007**, *129*, 2577 – 2587.
[40] R. Abel, T. Young, R. Farid, B. J. Berne, R. A. Friesner, *J. Am. Chem. Soc.* **2008**, *130*, 2817 – 2831.
[41] J. Michel, J. Tirado-Rives, W. L. Jorgensen, *J. Am. Chem. Soc.* **2009**, *131*, 15403 – 15411.
[42] D. D. Robinson, W. Sherman, R. Farid, *ChemMedChem* **2010**, *5*, 618 – 627.
[43] T. Beuming, Y. Che, R. Abel, B. Kim, V. Shanmugasundaram, W. Sherman, *Proteins Struct. Funct. Bioinf.* **2012**, *80*, 871 – 883.
[44] N. G. M. Davies, H. Browne, B. Davis, M. J. Drysdale, N. Foloppe, S. Geoffrey, B. Gibbons, T. Hart, R. Hubbard, M. R. Jensen, H. Mansell, A. Massey, N. Matassova, J. D. Moore, J. Murray, R. Pratt, S. Ray, A. Robertson, S. D. Roughley, J. Schoepfer, K. Scriven, H. Simmonite, S. Stokes, A. Surgenor, P. Webb, M. Wood, L. Wright, P. Brough, *Bioorg. Med. Chem.* **2012**, *20*, 6770 – 6789.
[45] S. Riniker, L. J. Barandun, F. Diederich, O. Krämer, A. Steffen, W. F. van Gunsteren, *J. Comput.-Aided Mol. Des.* **2012**, *26*, 1293 – 1309.
[46] K. Lafleur, D. Huang, T. Zhou, A. Caflisch, C. Nevado, *J. Med. Chem.* **2009**, *52*, 6433 – 6446.
[47] H. Zhao, J. Dong, K. Lafleur, C. Nevado, A. Caflisch, *ACS Med. Chem. Lett.* **2012**, *3*, 834 – 838.
[48] K. Lafleur, J. Dong, D. Huang, A. Caflisch, C. Nevado, *J. Med. Chem.* **2013**, *56*, 84 – 96.
[49] Conference notes on the lecture of P. Brennan, RSC Advances in Synthesis and Medicinal Chemistry, Welwyn Garden City, UK, May 1st, **2012**.
[50] R. M. Eglen, T. Reisine, P. Roby, N. Rouleau, C. Illy, R. Bossé, M. Bielefeld, *Curr. Chem. Genomics* **2008**, *1*, 2 – 10.
[51] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. F. Mayer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, *Arch. Biochem. Biophys.* **1978**, *185*, 584 – 591.
[52] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586 – 3616.
[53] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen, *J. Comput. Chem.* **2005**, *26*, 1701 – 1718.
[54] W. L. Jorgensen, J. Chandrasekhar, J. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926 – 935.
[55] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. MacKerell, Jr., *J. Comput. Chem.* **2010**, *31*, 671 – 690.
[56] T. Darden, D. York, L. G. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089.
[57] G. Bussi, D. Donadio, M. Parrinello, *J. Chem. Phys.* **2007**, *126*, 014101.
[58] B. R. Brooks, C. L. Brooks III, A. D. MacKerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, *J. Comput. Chem.* **2009**, *30*, 1545 – 1614.
[59] M. Seeber, M. Cecchini, F. Rao, G. Settanni, A. Caflisch, *Bioinformatics* **2007**, *23*, 2625.

Received: April 9, 2013
Published online on ■■ ▮▮, 0000

CHEMMEDCHEM Chemistry & Drug Discovery

Supporting Information

© Copyright Wiley-VCH Verlag GmbH & Co. KGaA, 69451 Weinheim, 2013

Structured Water Molecules in the Binding Site of Bromodomains Can Be Displaced by Cosolvent

Danzhi Huang,* Emanuele Rossini, Sandra Steiner, and Amedeo Caflisch*[a]

cmdc_201300156_sm_miscellaneous_information.pdf

Gromacs input parameters used in the simulations

```
integrator = md
dt = 0.002
nsteps = 500000000
nstcomm = 1
nstcalcenergy = 1
nstxout = 5000
nstvout = 5000
nstfout = 5000
nstlog = 5000
nstenergy = 5000
nstlist = 100
ns_type = grid
pbc = xyz
rlist = 1.0
coulombtype = PME
rcoulomb = 1.0
fourierspacing = 0.1
fourier_nx = 0
fourier_ny = 0
fourier_nz = 0
pme_order = 4
ewald_rtol = 1e-5
optimize_fft = yes
vdwtype = cutoff
rvdw = 1
Tcoupl = v-rescale
ld_seed = -1
tau_t = 0.1 0.1
```

```
tc-grps = Protein non-protein
ref_t = 310 310
; Pressure coupling is on
Pcoupl = berendsen
pcoupltype = isotropic
tau_p = 2
compressibility = 4.5e-5
ref_p = 1
gen_vel = no
gen_temp = 310
constraints = all-bonds
constraint_algorithm= lincs
```

Water positions	CREBBP	BAZ2B	ATOM	Distance (nm)
1	Ala1164	Cys1940	N	0.45
	Asn1163	Asn1939	O	0.45
	Tyr1125	Tyr1901	OH	0.35
2	Tyr1125	Tyr1901	OH	0.5
	Met1160	Val1936	O	0.35
	Ala1164	Cys1940	N	0.4
3	Met1133	Met1909	O	0.35
	Asp1134	Asp1910	CA	0.45
	Met1133	Met1909	N	0.6
4	Pro1110	Pro1888	O	0.35
	Gln1113	Gln1891	O	0.35
	Val1115	Val1893	N	0.45
5	Pro1110	Pro1888	O	0.35
	Pro1110	Pro1888	CB	0.45
	Val1115	Val1893	CA	0.6
6	Pro1110	Pro1888	O	0.4
	Gln1113	Gln1891	O	0.4
	Pro1114	Pro1892	O	0.4
7	Met1133	Met1909	N	0.35
	Tyr1125	Tyr1901	OH	0.35
	Met1133	Met1909	O	0.35

Table S1. Triplets of protein atoms used for calculation of occupancy. The distances listed in the column on the right are threshold values. The threshold values are decided based on the crystal structure of CREBBP (PDB code 3P1C). Protein atoms close to the corresponding crystal waters are selected to define their relative positions. Preference is given to non-hydrogen atoms which locate at protein backbone as they have lower flexibility than the side chain atoms. In case a selected atom can form a hydrogen bond with the corresponding water a threshold value 0.35 nm is applied, otherwise a larger threshold based on the original distance in the crystal structure is used. A position is considered occupied if all three distances are below the threshold values.

			Occupancy (%)	
Protein	Water position	Methanol	Ethanol	DMSO
BAZ2B	1	3.0	10.3	3.3
	2	0.2	0.3	0.0
	3	0.8	0.5	0.0
	4	0.1	0.1	0.0
	5	3.6	4.3	0.5
	6	0.2	0.3	0.0
	7	0.0	0.0	0.0
CREBBP	1	0.5	5.1	2.1
	2	0.2	0.2	0.0
	3	1.3	0.5	0.0
	4	0.0	0.0	0.0
	5	0.8	0.9	0.2
	6	0.3	0.9	0.0
	7	0.0	0.0	0.5

Table S2: Occupancies of the cosolvents at the seven water positions during 1 μs simulation. The same triplets of protein atoms and distance thresholds as for the water occupancy (Table S1) were used.

binding;unbinding	methanol			ethanol			DMSO		
thresholds (nm)	τ_{on} (ns)	τ_{off} (ns)	k_{off}/k_{on} (mM)	τ_{on} (ns)	τ_{off} (ns)	k_{off}/k_{on} (mM)	τ_{on} (ns)	τ_{off} (ns)	k_{off}/k_{on} (mM)
BAZ2B									
0.3;0.7	1.2	1.3	431	1.2	3.5	148	2.7	6.7	179
0.4;0.7	1.0	0.8	603	1.1	2.6	184	2.5	5.1	212
0.5;0.7	0.8	0.7	560	0.7	2.0	158	2.2	3.8	250
0.3;0.8	1.2	1.8	301	1.2	5.3	102	2.6	9.5	118
0.4;0.8	**1.1**	**0.9**	**543**	**0.7**	**2.4**	**124**	**2.6**	**8.9**	**128**
0.5;0.8	0.9	0.8	475	0.8	2.7	129	2.3	6.9	148
0.3;0.9	1.2	2.7	194	1.2	6.7	79	2.4	10.9	95
0.4;0.9	1.0	1.2	397	1.1	5.5	85	2.7	10.5	111
0.5;0.9	0.9	1.0	421	0.8	3.7	97	2.3	9.6	104
CREBBP									
0.3;0.7	1.3	1.3	439	1.1	4.3	108	1.9	25.2	32
0.4;0.7	1.2	1.1	482	0.9	3.9	106	2.1	23.6	39
0.5;0.7	1.0	0.8	551	0.9	3.2	118	2.1	21.7	42
0.3;0.8	1.4	1.3	463	1.1	4.8	98	1.8	28.2	28
0.4;0.8	**0.5**	**1.3**	**176**	**1.0**	**4.5**	**94**	**2.1**	**26.0**	**36**
0.5;0.8	1.1	1.1	436	0.9	3.9	97	2.1	25.0	36
0.3;0.9	1.4	1.5	413	1.1	6.7	70	1.7	30.2	24
0.4;0.9	1.3	1.5	393	0.9	5.8	71	2.0	30.4	28
0.5;0.9	1.1	1.3	384	0.9	4.8	77	1.8	27.8	28

Table S3: **Robustness analysis.** Dissociation constant from fitting of cumulative distributions of unbinding and binding times using different thresholds. Values in boldface are reported in Table 1 of the main text. The characteristic time of the slow phase of the double-exponential fitting is used to calculate the binding rate $k_{on} = 1/(\tau_{on}[\text{cosolvent}])$ and unbinding rate $k_{off} = 1/\tau_{off}$. The concentration of the cosolvent in the simulation box is 440 mM.

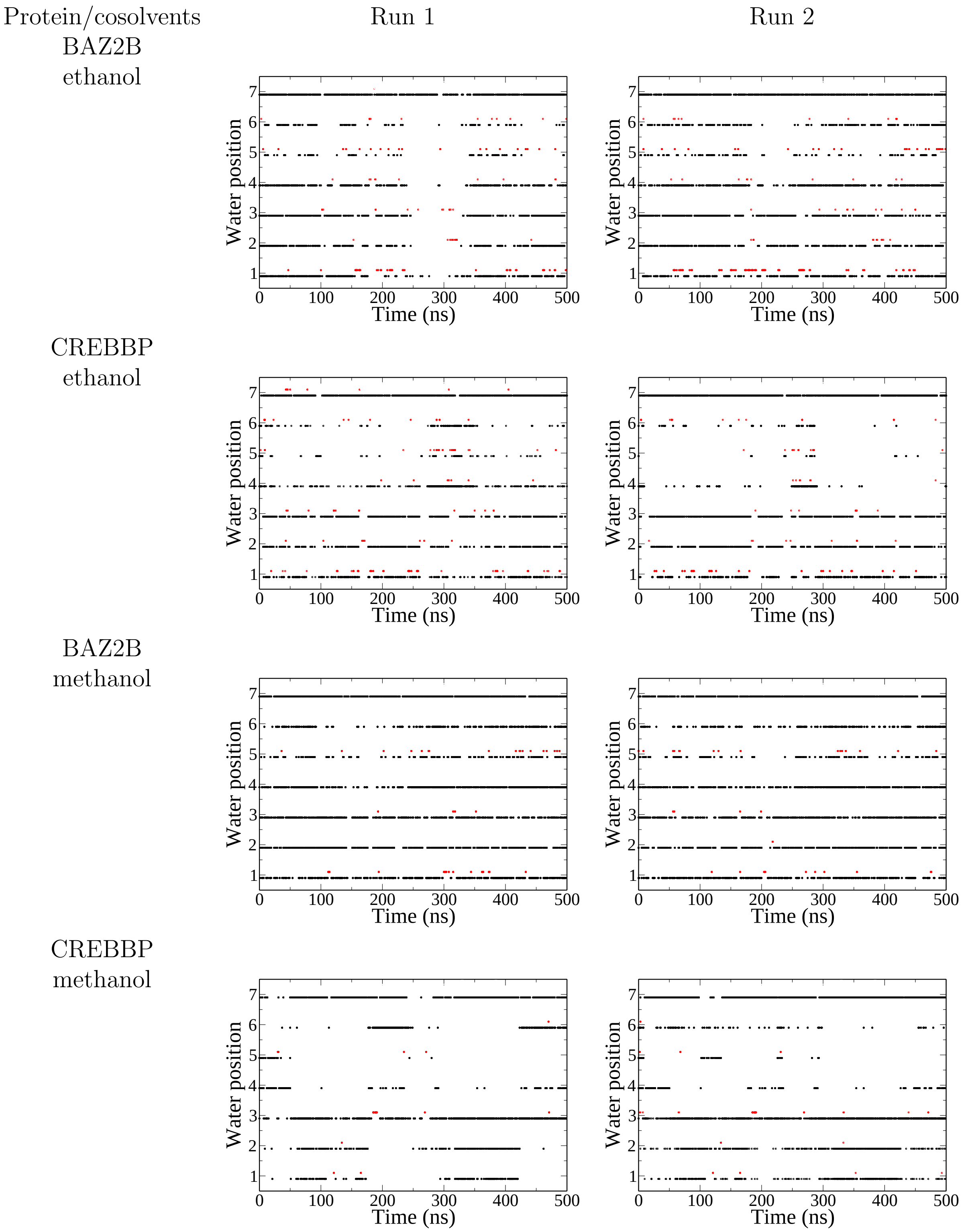
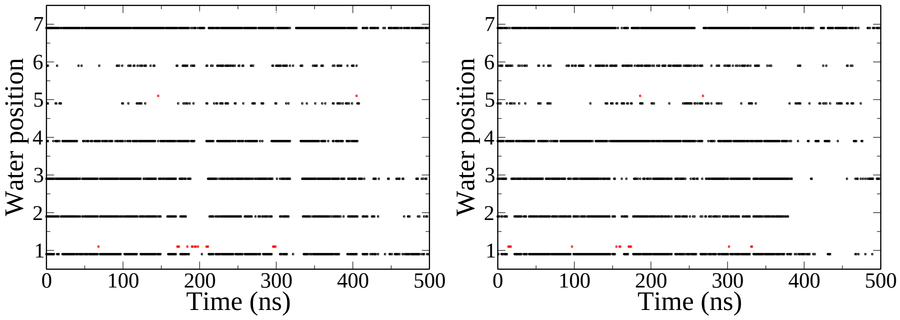


Figure S1: Caption on next side.

BAZ2B
DMSO



CREBBP
DMSO

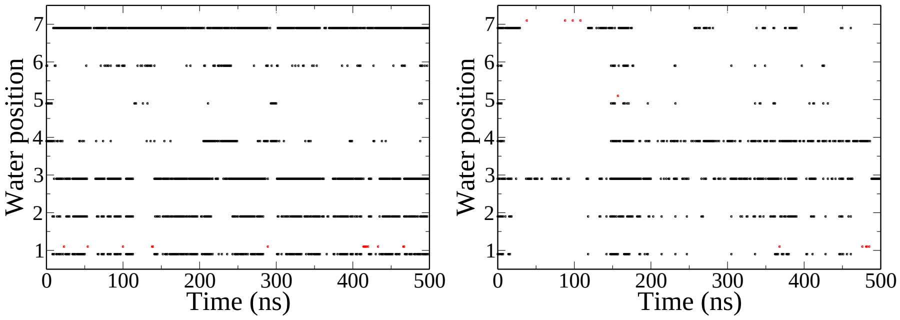


Figure S1. Cosolvent occupancy at the positions of the structured water molecules. The time series show the presence of water (black dot) or cosolvent (red dot) along time intervals of 1 ns.

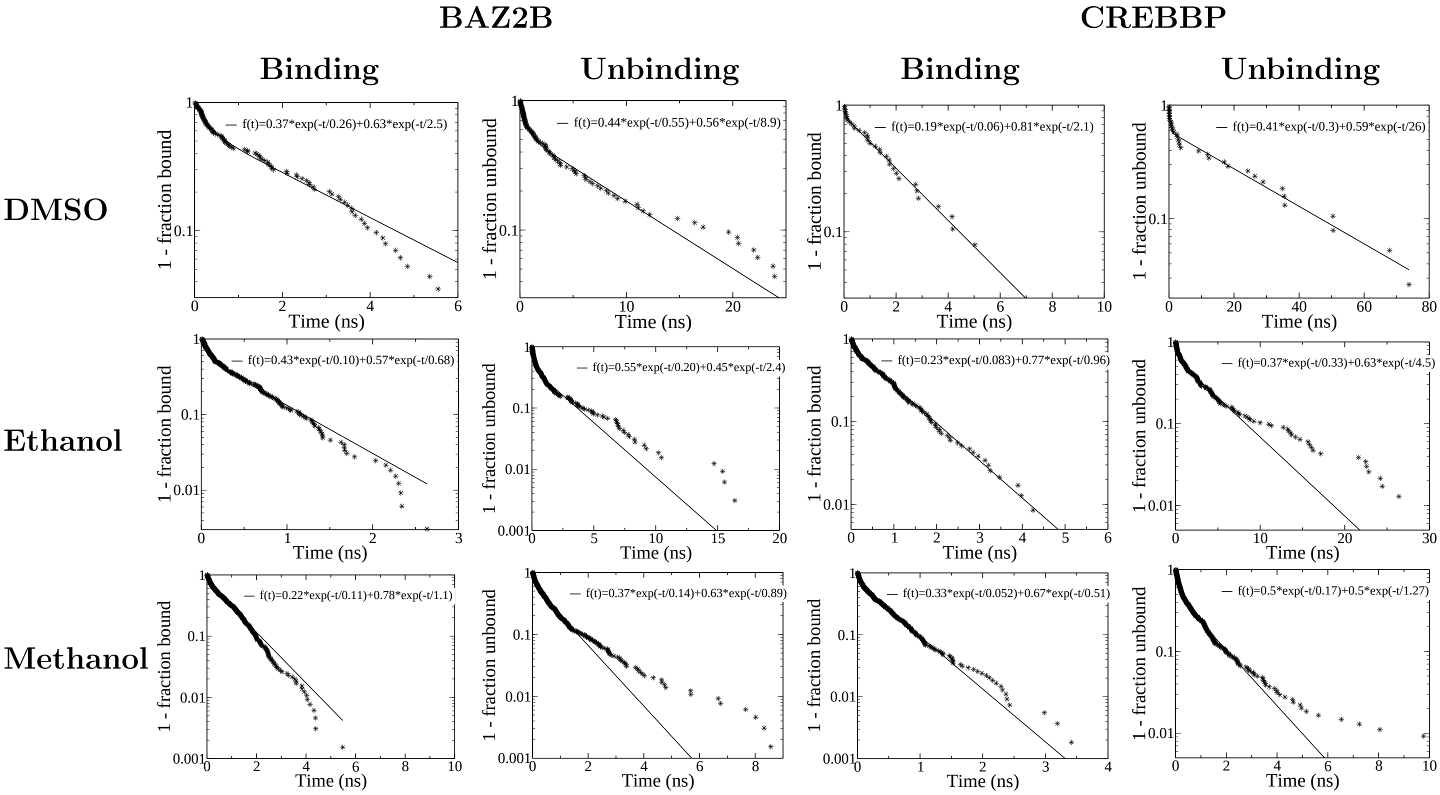


Figure S2: Cumulative distribution of the binding time and unbinding time f(t) $= \int_t^\infty p(\tau) d\tau$, where p is the probability distribution of the binding and unbinding time, respectively. Binding and unbinding events are defined by a separation between the centers of mass of the acetyl-lysine binding site and cosolvents smaller than 0.4 nm and larger than 0.8 nm, respectively. These threshold values were chosen upon visual analysis of the time series. The stars represent the binding and unbinding events observed in the two MD runs. The solid lines are double-exponential fits.

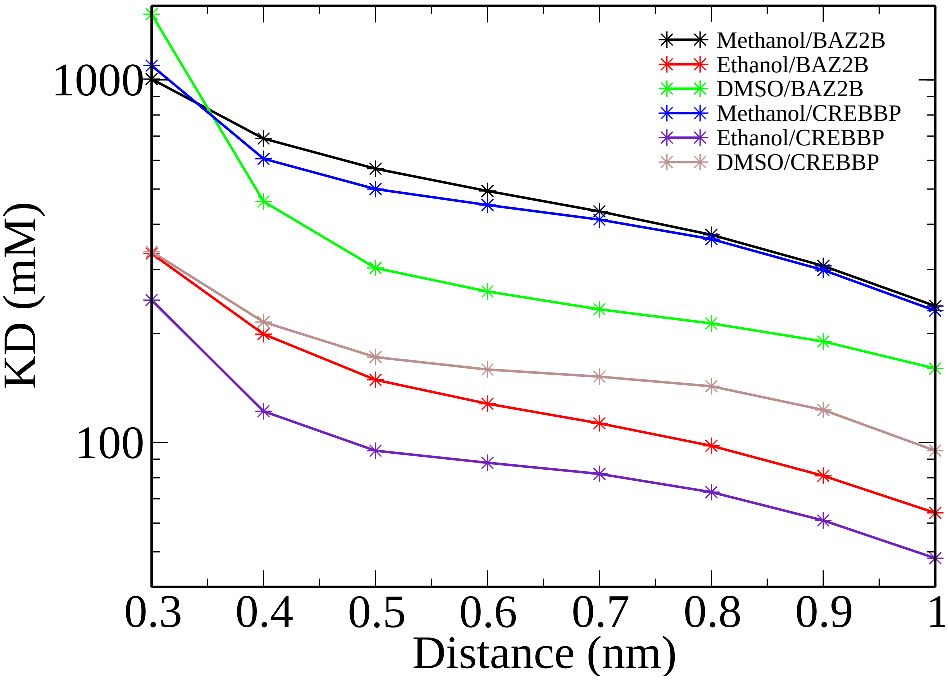


Figure S3: **Robustness of dissociation constant upon changes of occupancy threshold.** Values of the threshold ranging from 0.3 to 1.0 nm are used to calculated the occupancy of the acetyl-lysine binding site. The dissociation constant is calculated from the occupancy using K_D = [cosolvent] [unbound protein] / [bound protein] = [cosolvent] (100 - occupancy) / occupancy.

Chapter 7

Mechanism and kinetics of acetyl-lysine binding to bromodomains

Steiner, S.[‡], Magno, A.[‡] and Caflisch, A.; *Journal of Chemical Theory and Computation*, 9(9): 4225-4232, 2013

[‡]These authors contributed equally to this work

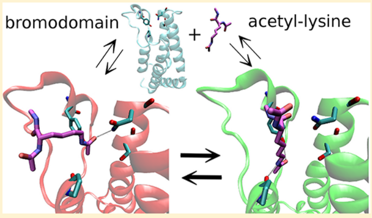
JCTC Journal of Chemical Theory and Computation
Article
pubs.acs.org/JCTC

Mechanism and Kinetics of Acetyl-Lysine Binding to Bromodomains

A. Magno,[†,‡] S. Steiner,[†,‡] and A. Caflisch*[,†]

[†]Department of Biochemistry, University of Zurich, Winterthurerstrasse 190 CH-8057 Zurich, Switzerland

S *Supporting Information*



ABSTRACT: Bromodomains are four-helix bundle proteins that specifically recognize acetylation of lysine side chains on histones. The available X-ray structures of bromodomain/histone tail complexes show that the conserved Asn residue in the loop between helices B and C is involved in a hydrogen bond with the acetyl-lysine side chain. Here we analyze the spontaneous binding of acetyl-lysine to the bromodomain TAF1(2) by the first molecular dynamics simulations of histone mark binding to an epigenetic reader protein. Multiple events of reversible association sampled along the unbiased simulations allow us to determine the pathway and kinetics of binding. The simulations show that acetyl-lysine has two major binding modes in TAF1(2) one of which corresponds to the available crystal structures and is stabilized by a hydrogen bond to the conserved Asn side chain. The other major binding mode is more buried than in the crystal structures and is stabilized by two hydrogen bonds with conserved residues of the loop between helices Z and A. In the more buried binding conformation, three of the six structured water molecules at the bottom of the binding pocket are displaced by the acetyl-lysine side chain. The kinetic analysis shows that the two binding modes interconvert on a faster time scale with respect to the association/dissociation process. The atomic-level description of the binding pathway and binding modes is useful for the design of small molecule modulators of histone binding to bromodomains.

INTRODUCTION

Bromodomains are protein modules of about 110 residues which recognize acetylated lysine side chains mainly in histones and are thus involved in transcriptional regulation.[1,2] In the human genome 46 proteins with a total of 61 different bromodomains have been identified, with up to 6 bromodomains per protein.[3] Bromodomains occur often in tandem with a PHD module or with other bromodomain(s), suggesting that combinatorial readout may be at work in translating the histone code.[2,4,5] The potential role of bromodomains in tumors and inflammation[4,6,7] has spurred large-scale structural studies with the ultimate goal to facilitate the discovery of small-molecule inhibitors able to interfere in the process of reading acetylated lysine.[8,9] In 1999, the first three-dimensional structure of a human bromodomain was solved;[10] since then, the crystal and/or solution structures of more than 40 human bromodomains have been reported.[11,12]

All available structures show a conserved four-helix bundle topology in which the ZA-loop and BC-loop connect the first two α helices (called Z and A) and last two α helices (called B and C), respectively.[12,13] The acetyl-lysine binding site is very similar in all structures of bromodomains.[12] There are several crystal structures with bromodomains in complex with acetyl-lysine[4] (e.g., PDB codes 3P1C, 3O34, and 3O35) or peptides with two acetylated lysine side chains (PDB codes 3UVW, 3UVX, 3UVY, and 3UW9). These structures share a common intermolecular hydrogen bond between the side chain of the conserved Asn in the BC loop and the acetyl oxygen of the ligand. The binding mode is further stabilized by a network of water-mediated hydrogen bonds at the bottom of the binding groove and hydrophobic contacts.[14] In addition to the structural characteristics, thermodynamic parameters of histone tail binding have been measured by isothermal titration calorimetry using mono- and diacetylated histone peptide segments.[4] The measured dissociation constants are in the high micromolar range, and seem rather insensitive to the peptide length. In general, the dynamical character of the process of ligand binding makes it a challenging investigation subject.[15] Molecular dynamics (MD) simulations with explicit solvent provide a useful tool to study binding of small molecules to proteins at the atomic level of detail.[15−22]

Here, we have employed MD simulations in conjunction with cut-based free energy profile analysis[15,23,24] to shed light on the spontaneous association of acetyl-lysine to the second bromodomain of the human transcription initiation factor TFIID subunit 1 (abbreviated as TAF1(2)). This bromodomain was chosen based on a previous study[25] in which the side chain of the conserved Asn showed two interconverting orientations. In particular, the side chain amide nitrogen of TAF1(2) Asn1604 populates a solvent exposed orientation which is rotated by about 180 degrees with respect to the orientation observed in most crystal structures. In this context there are several interesting questions which can be addressed by simulations: Is conventional MD able to reproduce the experimentally observed binding mode of the acetyl-lysine side chain? On which time scale does spontaneous binding take place? Which role does the side chain orientation of the conserved Asn play in the binding process? Are there alternative binding modes? The simulations describe with

Received: May 3, 2013
Published: July 10, 2013

© 2013 American Chemical Society

dx.doi.org/10.1021/ct400361k | *J. Chem. Theory Comput.* 2013, 9, 4225−4232

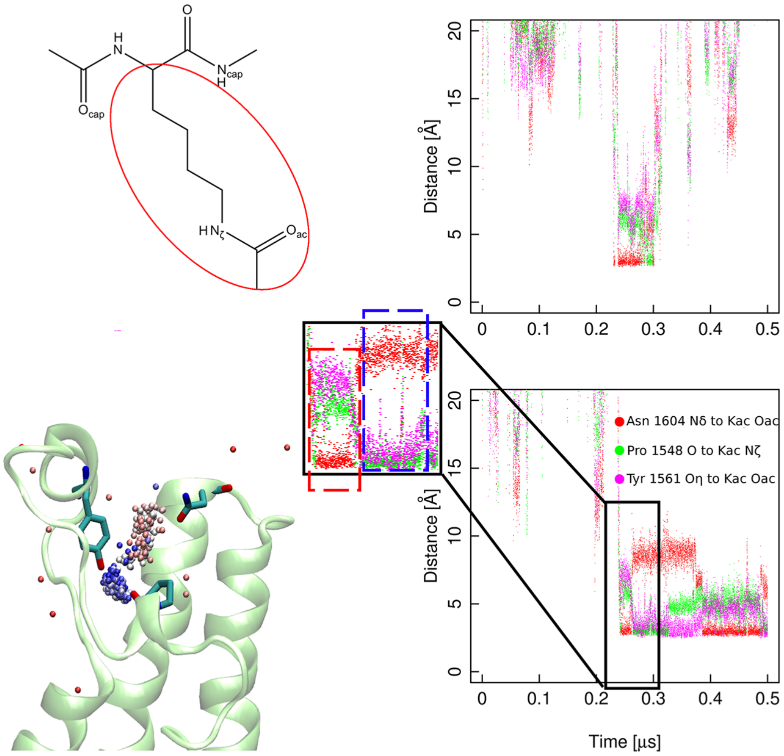


Figure 1. Spontaneous binding of Kac to TAF1(2). (top left) Two-dimensional structure of Kac. The red oval highlights the Kac side chain. (bottom left) Spheres representing the position of the carbonyl carbon atom of the Kac side chain. Their color reflects the simulation time from red to blue. The backbone of TAF1(2) and the side chains involved in binding are shown by a green ribbon and sticks, respectively. (right) Time series of key interactions along two independent simulations. The inset emphasizes the binding event shown in the bottom left panel and in Supporting Information Movie S1. The red and blue dashed rectangles in the inset illustrate the two most populated binding modes and correspond to the colors of the spheres in the bottom left panel.

atomistic resolution acetyl-lysine binding to the position observed in the crystal structure. Moreover, they reveal an alternative binding mode and provide kinetic information on the binding/unbinding processes.

RESULTS AND DISCUSSION

Twenty-four independent simulations of 0.5 μs each were started with acetylated lysine (with neutral blocking groups at its termini, abbreviated as Kac, see Figure 1 for the 2D structure) positioned randomly at a distance of more than 20 Å from the binding site of TAF1(2). No artificial guiding force or bias toward the binding site was used. In 16 of the 24 MD runs at least one event of spontaneous binding to the acetyl-lysine pocket was observed. In the remaining 8 MD runs Kac did not reach the binding site and remained most of the time in the solvent or was sporadically involved in unspecific association to different regions of the TAF1(2) surface. Unless stated otherwise, the following analysis refers to the 8-μs sampling of the 16 runs in which binding events were observed (a total of 4×10^5 snapshots).

Spontaneous Binding. The time series of some of the key interactions show that spontaneous and reversible binding of Kac to the acetyl-lysine pocket can take place in the 0.5 μs time scale of the MD runs (Figures 1 and 2 and Supporting Information Figures S1, S2, and movie files, ct400361k_si_001.mpg (Movie S1) and ct400361k_si_002.mpg (Movie S2)). From the analysis of hydrogen bond frequencies, it emerges that the hydrogen bond between the acetyl oxygen of the Kac side chain and the side chain of the conserved Asn1604 in the BC loop is present in nearly 40% of the snapshots of the bound state. Interestingly,

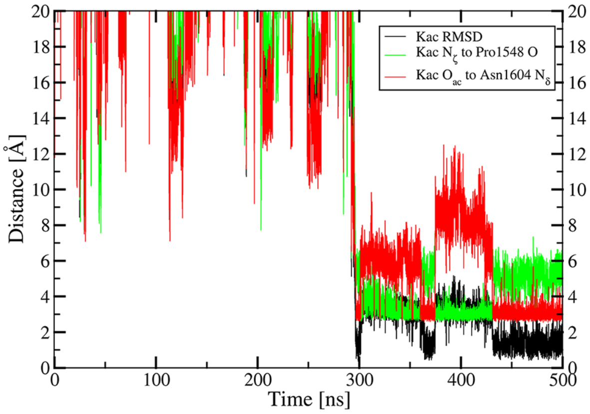


Figure 2. Spontaneous binding to the crystallographic binding mode. The plot shows the time series of the RMSD from the X-ray structure (black) together with the distances that characterize the hydrogen bond observed in the crystal structures (red) and the hydrogen bond of the more buried binding mode revealed by the present MD study (green). Note the frequent interconversions between the two binding modes. The RMSD of the Kac side chain atoms was calculated upon optimal overlap of the C_α atoms in the four helices. The reference structure for the RMSD calculation is the crystal structure of CREBBP in complex with Kac (PDB code 3P1C) as there is no crystal structure of TAF1(2) with Kac or a histone tail peptide. The MD run used for this plot is shown in Supporting Information Movie S2.

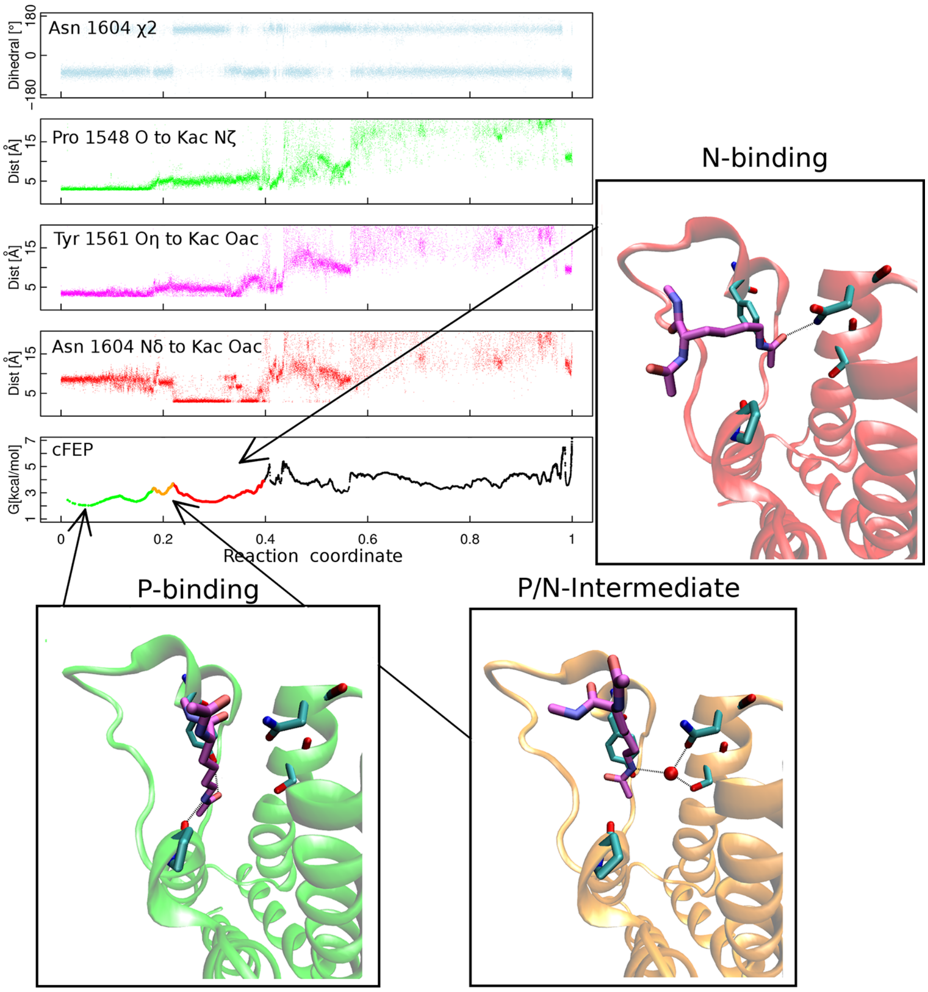


Figure 3. Geometric variables used to analyze binding (top) and cut-based free energy profile (abbreviated as cFEP) calculated using the mean first passage time from the representative node of the P-binding mode (bottom). The reaction coordinate of the cFEP is the relative partition function (see the Methods section). The colors of the cFEP reflect the three main basins whose representatives are shown with TAF1(2) in ribbon model and Kac in sticks (insets). The unbound state is in black in the cFEP. The 400 000 MD snapshots were ordered according to the free-energy sub-basin they belong to. In this way, each of the sub-basins on the free energy profile is annotated with the corresponding values of the geometric variables. As an example, the first sub-basin on the left is the P-binding mode which has hydrogen bond distance for Pro 1548 O—Kac N_ζ and Tyr1561 O_η—Kac O_{ac} whereas the Asn1604 N_δ—Kac O_{ac} distance is about 8–9 Å.

two residues of the ZA loop are also involved in hydrogen bonds with Kac. The backbone oxygen of Pro1548 (in the so-called WPF shelf) and the side chain hydroxyl of Tyr1561 act frequently as hydrogen bond acceptor and donor, respectively.

Two Major Binding Modes. The trajectories with spontaneous events of association and the corresponding time series of key intermolecular distances indicate that there are two distinct binding modes (red and blue spheres, respectively, in Figure 1 bottom, left). From the unbound state, Kac adopts first a binding mode matching the one observed in several crystal structures with the hydrogen bond between the Asn1604 N_δ atom and the Kac side chain oxygen O_{ac} (called N-binding mode hereafter). The time series of the Kac root-mean-square deviation (RMSD) from the crystallographic binding mode shows several simulation segments during which the RMSD oscillates between 0.5 and 1.5 Å (e.g., the last 60 ns of the MD run shown in Figure 2) which confirms that the N-binding mode corresponds to the binding mode determined experimentally. Note that the structure of TAF1(2) does not change upon Kac binding as monitored by the RMSD of the C_α atoms from the structure at the beginning of the MD run (see Supporting Information Figure S3) and also shown in Supporting Information Movie S2.

From the N-binding mode, Kac can reach reversibly a binding conformation which is more buried than in the crystal structure and in which the oxygen atom of Pro1548 acts as hydrogen bond acceptor for the Kac side chain N_ζ (called P-binding mode hereafter). Further stabilization is provided by a hydrogen bond between the hydroxyl group of Tyr1561 and the Kac side chain O_{ac}, as well as between atom N_δ of Asn1604 and either of the two oxygen atoms in the backbone moiety of Kac, i.e., O_{cap} (N-terminal capping group) or O (C-terminal group). An MD run with a spontaneous binding event and further interconversions between N- and P-binding modes is shown in Figure 1 and Supporting Information Movie S2.

Clustering of the snapshots based on pairwise RMSD of the Kac side chain atoms and consecutive calculation of the cut-based free energy profile[23] confirm the existence of two major binding modes (Figure 3). It is important to note that the P-binding mode is more buried (Figure 1 bottom, left) and slightly more populated than the N-binding mode. The former includes about 50% of the MD snapshots in the bound state while the latter includes about 40%. Since the P-binding mode has not been reported as of today, we decided to run additional simulations of Kac binding to TAF1(2) with a force field different from CHARMM27.[26] The N- and P-binding modes are also observed in each of four 1 μs runs with the AMBER[27] force field, and they show similar statistical weight (Supporting Information Figure S4).

We also wanted to exclude the idea that the more buried P-binding mode is an artifact due to the simplified model of the ligand. For this purpose, we run eight additional simulations of 1 μs each with the Kac-Gly-Gly-Kac tetrapeptide (derived from the diacetylated histone tail H4K5/K8) using two different bromodomains (i.e., CREBBP and BRD4(1)) and with two different force fields (CHARMM27[26] and AMBER,[27] Supporting Information Table S1). The P-binding mode is populated also by the tetrapeptide ligand in each of the two bromodomains and irrespective of the force field used (Supporting Information Figure S4). Thus, the results of the control simulations provide evidence that the P-binding mode is not an artifact of the force field or the simplified ligand. It is also important to note that in the N-binding mode the RMSD from the crystal structure of the side chain of the N-terminal Kac is almost always below 2 Å (Supporting Information Figure S5). Moreover, the C-terminal Kac, i.e., the acetyl-lysine side chain that is not in contact with the conserved Asn, has a more pronounced flexibility than the N-terminal Kac (Supporting Information Figure S5) which is consistent with the larger temperature factors of the C-terminal than the N-terminal acetyl-lysine in the crystal structure of the complex with BRD4(1) (PDB code 3UVW).

Kinetics. The existence of two major binding modes calls for further investigations of the time scales required for binding/unbinding and how they relate to the transitions within the bound state. For this analysis, first individual free energy basins were isolated by the cut-based free energy profile method following an iterative procedure reported previously[24] (see also the Methods section). Besides the two major binding modes, there is a metastable state (called P/N-intermediate and populated at about 10%) formed by MD snapshots in which the side chain of Kac points inside the binding groove but is arranged such that direct hydrogen bonds with Pro1548, Tyr1561, or Asn1604 are not possible. This binding conformation is stabilized mainly by a water-bridged hydrogen bond between Kac N_ζ and Asn1604 O_δ (see bottom, right inset in Figure 3).

The mean first passage time (MFPT) between free energy basins was evaluated using representative clusters selected according to statistical weight as published previously.[24] The MFPT values indicate that there are two well-separated time scales (Table 1). While the two major binding modes interconvert relatively fast, i.e., with an MFPT of ≈20 ns, binding/unbinding takes place on a much slower time scale (MFPT ≈ 300 ns). Figure 4 illustrates these two kinetic regimes by the free energy profiles plotted together with the MFPT to the reference state. The free energy barrier to exit the bound state (located at a value of the reaction coordinate of about 0.4 in Figure 4) is significantly higher than the barriers between the three basins in the bound state. The number of direct transitions between basins is shown in Table 1. The large number of direct transitions between P-binding and N-binding modes compared to the much smaller number of binding/unbinding events reflects the kinetic separation of the two processes which emerges also from the MFPT calculations. It is important to note that the number of binding and unbinding events (20 and 10, respectively) determined using the definition of basins by the cut-based free energy profile agrees with the visual inspection of the trajectories and the time series presented in Supporting Information Figures S1 and S2. Such agreement provides strong evidence that the clustering and the definition of free energy basins by the cut-based free energy profile method capture the relevant features of the binding and unbinding processes.

Table 1. Kinetics Data Extracted from the Simulations[a]

transitions (MFPT [ns]) from\to	unbound	P-binding	N-binding	P/N-intermediate
unbound		1 (220)	14 (220)	5 (220)
P-binding	2 (320)		129 (28)	310 (20)
N-binding	4 (320)	135 (15)		17 (29)
P/N-intermediate	4 (320)	312 (6)	16 (30)	

[a] The table shows the number of direct transitions between pairs of states as observed in the MD simulations. The MFPT values in nanoseconds are given in parentheses. Note that the MFPT values are calculated from the transition matrix, i.e., they take into account both direct and indirect transitions.

Dissociation Constant. The multiple events of spontaneous and reversible binding allow one to calculate the dissociation constant from the ratio of the off-rate and on-rate. The off-rate is the inverse of the MFPT from any of the binding modes to the unbound state (i.e., 1/320 ns^{-1}; see Table 1). The on-rate is the inverse of the MFPT to any of the bound states multiplied by the concentration of Kac in the simulation box (1/(220 ns × 6 mM)). It thus follows that the dissociation constant is about $(220 \times 6)/320 \simeq 4$ mM. This value is about 1 order of magnitude higher than the dissociation constants of mono- and diacetylated histone peptides as measured by isothermal titration calorimetry[4] which is probably a consequence of the smaller ligand used in the simulations than in the experiments.

Orientation of the Side Chain Amide Group of the Conserved Asn. The preferred orientation of the side chain of the conserved Asn1604 differs from the one observed in most bromodomain crystal structures, i.e., the N_δ atom points preferentially toward the solvent ($\chi_2 \approx -70°$, Figure 3 top) rather than the bottom of the binding site ($\chi_2 \approx 120°$). The solvent exposed orientation of the Asn1604 $-NH_2$ group is stabilized by a water-bridged hydrogen bond between the Asn side chain carbonyl and the side chain hydroxyl of Ser1600 in the C-terminal turn of helix B. This observation raises the question whether different binding modes of Kac prefer different orientations of the Asn1604 side chain. Here, it is important to note that based on a previous work[25] half of the simulation runs were started with the solvent exposed orientation of the Asn1604 —NH_2 group and the other half with its orientation toward the binding site (see Methods). Figure 5 shows the histograms of the Asn1604 χ_2 dihedral values computed over the snapshots belonging to the three individual binding modes as well as for the segments of the trajectories during which the ligand is unbound. Note, that for the unbound conformation the histogram is characterized by two peaks corresponding to the two different orientations with a preference for the solvent exposed orientation. Importantly,

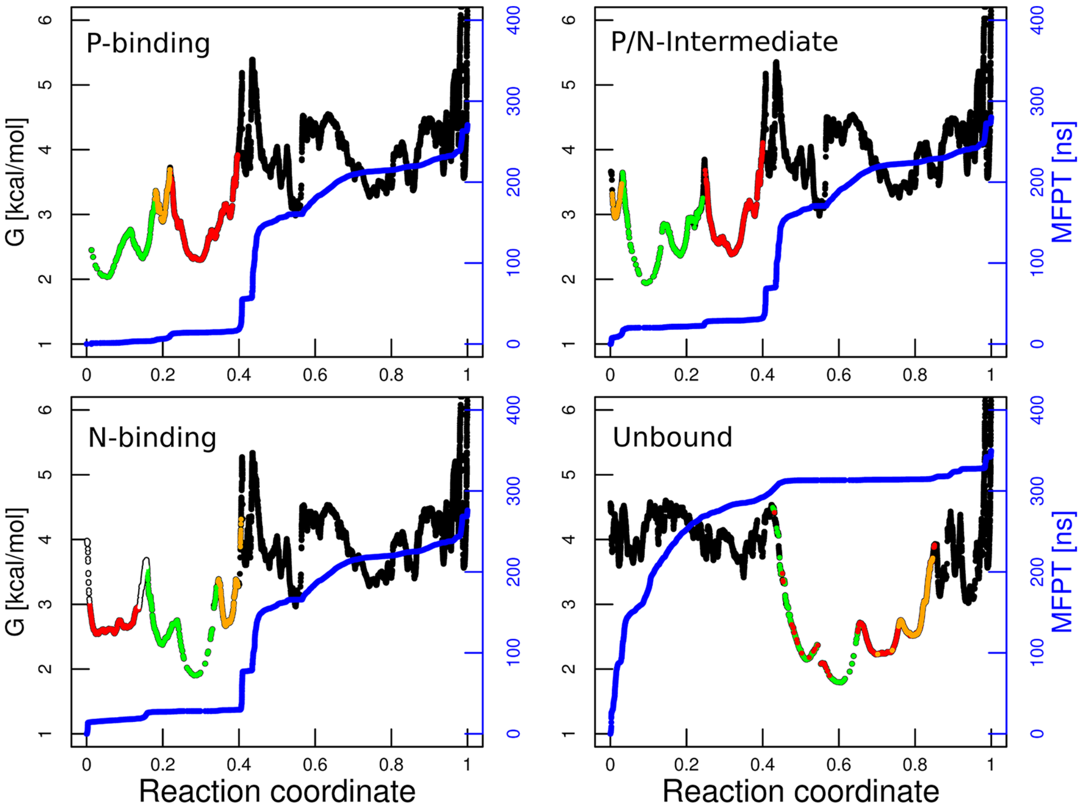


Figure 4. Four panels show the cut-based free energy profiles with the three equilibrium bound conformations or the unbound state as reference. Each panel is labeled with the state used as reference which is the first basin on the left. Individual basins are colored according to the same color scheme as in Figure 3, i.e., green, red, orange, and black for P-binding mode, N-binding mode, P/N-intermediate, and unbound state, respectively. The empty circles in the bottom left panel represent the nodes on the barrier between the N-basin and the unbound state. The reaction coordinate is the relative partition function (see the Methods section). The MFPT values to the representative node of the reference basin are also shown (blue lines with y-axis on the right). The MFPT is a monotonous growing function because the nodes are sorted according to their kinetic distance from the reference basin. Note that the three binding modes have approximately equal kinetic separation from the unbound conformation (i.e., 320 ns; see Table 1). Thus, there is a significant overlap of the three binding modes when plotting the profile from a node belonging to the unbound conformation (bottom right panel).

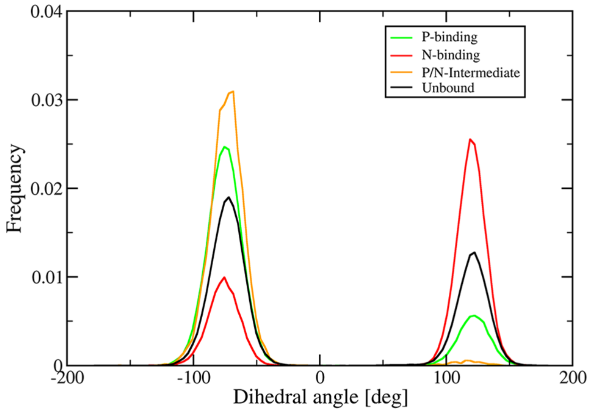


Figure 5. Histograms of the two main conformations of the χ_2 angle of the conserved Asn1604 for the individual binding modes (colors) and the unbound state (black). Positive values of the χ_2 angle correspond to the orientation of Asn1604 N_δ toward the bottom of the binding site while negative values to the orientation of Asn1604 N_δ toward the solvent.

as experimentally observed in several X-ray structures of the acetyl-lysine−bromodomain complex, the N-binding conformation peaks at $\chi_2 \approx 120°$, i.e., the orientation of Asn1604 —NH_2 pointing toward the binding site. On the contrary, in the P-binding and the P/N-intermediate modes the orientation of the —NH_2 group of Asn1604 toward the solvent is preferred, since in the former it frequently acts as donor for the O or O_{cap} atoms of Kac, while in the latter a water molecule coordinates O_δ of Asn1604, O_γ of Ser1600, and N_ζ of Kac (see Figure 3). Furthermore, it is interesting to analyze the role of Asn1604 during the binding process. For this purpose, the N-binding mode is decomposed into two sub-basins made up of snapshots with positive and negative values, respectively, of the Asn1604 χ_2 dihedral. Strikingly, 12 out of the 14 transitions from the unbound state to the N-binding state take place when the N_δ atom of the Asn1604 points toward the solvent (Supporting Information Table S2 and Movie S1). The Kac ligand can then further penetrate the binding groove, ending up in one of the major binding modes.

Conserved Water Molecules in the Acetyl-Lysine Binding Site. Six water molecules stabilized by hydrogen bonds with polar groups in the bottom of the acetyl-lysine binding site seem to be conserved in most crystal structures of bromodomains.[8] In the N-binding mode, the six conserved water molecules are arranged in the same fashion as observed in the crystal structures. In contrast, in the P-binding mode, water molecules 2, 3, and 4 (nomenclature of Figure 1 of ref 8) are replaced by the Kac side chain which, as mentioned above, is more buried than in the N-binding mode (Figure 6). This simulation result is useful for the design of small molecule modulators of histone binding to bromodomains. In particular, some of the conserved water molecules need not necessarily be present in high-throughput docking campaigns. Moreover, optimization of known binders could focus on the replacement of one or more of the structured water molecules by functional groups with similar hydrogen bonding properties.

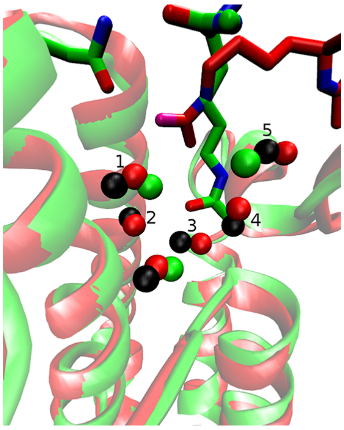


Figure 6. Structured water molecules are conserved in the N-binding mode while three of them are displaced by the tip of the Kac side chain in the P-binding mode. The representative snapshots of the N-binding mode (red) and P-binding mode (green) are aligned structurally to the crystal structure of apo TAF1(2) (PDB code 3UV4) using only the C_α atoms of the bromodomain helices. The bromodomain backbone, Kac, and structured water molecules are shown (ribbons, sticks, and spheres, respectively) for the two binding modes. For the crystal structure, only the structured water molecules are shown (black spheres) to avoid overcrowding of the figure. The Asn1604 side chain of the P-binding mode representative is also shown (carbon atoms in green). The labeling of the structured water molecules is the same as in ref 8. Water molecules 2−4 are displaced in the P-binding mode.

■ CONCLUSIONS

We have analyzed the pathway and kinetics of Kac binding to the TAF1(2) bromodomain by multiple MD simulations of spontaneous and reversible binding. The MD trajectories reveal two major binding modes of Kac which differ in the bromodomain residues providing the anchoring interactions. From the unbound state, Kac reaches first a position in which it is stabilized by a hydrogen bond between its side chain carbonyl oxygen and the side chain amide of the conserved Asn. This binding mode is identical to the binding mode observed in several crystal structures of bromodomains with Kac or short peptide segments from histone tails. From the crystallographic binding mode Kac further penetrates the binding site and reaches a more buried binding mode in which the Kac side chain is involved in a hydrogen bond with the carbonyl group of the conserved Pro in the so-called WPF shelf. The WPF shelf is a highly conserved motif of the bromodomain sequence and has been highlighted as an important hotspot to increase binding affinity of small molecule inhibitors.[28,29] Moreover, the binding of diacetylated histone tails involves the WPF shelf which interacts with the acetyl-lysine side chain that is not in contact with the conserved Asn.[4,30] Furthermore, the six structured water molecules observed in most crystal structures are present in the crystallographic binding conformation, while three of them are replaced by the Kac side chain in the more buried binding mode. Importantly, in several crystal structures of bromodomains in complex with inhibitors (e.g., PDB codes 4HXK, 4HXO, 4IR3, 4A9I, and 4ALC) the carbonyl oxygen of the Pro of the WPF shelf is involved in a polar interaction with the inhibitor, e.g., with an aromatic CH of the [1,2,4]triazolo[4,3-*a*]pyridine moiety in the complex structure with BRD4(1) (PDB code 4HXO).[31] The crystal structure of the complex between BRD4(1) and 6,7-dihydrothieno[3,2-*c*]pyridin-5(4*H*)-yl(1*H*-imidazol-1-yl)methanone (PDB code 4HXK) is of particular interest as the 6,7-dihydrothieno[3,2-c]pyridin moiety is almost as buried as acetyl-lysine in the P-binding mode (Supporting Information Figure S6). Moreover, one of the structured water molecules is displaced by the dihydrothieno ring.[31]

The two binding modes observed in the MD simulations are distinct and separated by a free energy barrier which is smaller than the barrier of unbinding. In other words, upon binding there are several transitions between the two binding modes which can interconvert about 1 order of magnitude faster than the time required for full dissociation which is about 300 ns. The simulation methods used here, unbiased MD with explicit solvent (carried out on conventional compute clusters) and cut-based free energy profile analysis, can be applied to determine the binding mode(s) and kinetics of small molecules to proteins thus providing dynamical information which is useful to complement the static picture obtained by conventional experimental techniques. Although the histone code may consist of multiple post-translational modifications, the mechanism and kinetics of acetyl-lysine binding to bromodomains is a first step toward the in-depth understanding of the role of bromodomains in transcriptional regulation.

■ METHODS

Simulation Protocols. The coordinates of the TAF1(2) bromodomain (PDB code 3UV4) were downloaded from the protein database (URL www.pdb.org).[32] To reproduce neutral pH conditions, the side chains of aspartates and glutamates were negatively charged, those of lysines and arginines were positively charged, and the histidine side chains were neutral. Thirteen crystal water molecules inside or close to the binding site were kept while the remaining water molecules were deleted. Subsequently, the structure was solvated in a water box whose size was chosen to have a minimal distance of 12 Å between the boundary and any atom of the protein. The simulation system contained sodium and chloride ions to approximate an ionic strength of about 150 mM and to compensate for the total charge of the bromodomain. Preliminary MD simulations were carried out with CHARMM,[33] while the production runs were carried out with GROMACS 4.5.4[34] using the CHARMM PARAM27 all atom force field[26,35] and the TIP3P model of water.[36]

Periodic boundary conditions were applied, and electrostatic interactions were evaluated using the particle-mesh Ewald summation method.[37] The van der Waals interactions were truncated at a cutoff of 10 Å. The temperature of 310 K was kept constant by an external bath with velocity rescaling,[38] and the pressure was kept close to 1 atm by the Berendsen barostat.[39] The LINCS algorithm was used to fix the covalent bonds involving hydrogen atoms. The integration time step was 2 fs, and snapshots were saved every 20 ps.

Multiple independent MD runs of spontaneous binding of Kac to TAF1(2) were carried out. To mimick the context of the longer histone sequence, the N-terminus and C-terminus of Kac were neutralized by capping with acetyl and *N*-methyl-amide, respectively (Figure 1). Two different starting structures were generated for TAF1(2) by swapping the coordinates of Asn1604 O_δ and N_δ atoms in the X-ray structure (before hydrogen generation). Thus two different χ_2 dihedrals of Asn1604 were represented. Six starting positions for the Kac molecule were generated by randomly placing Kac with its center of mass ≈20 Å away from the center of mass of Asn1604 using the package PyMOL.[40] For each of the six Kac starting positions and each of the two TAF1(2) starting structures, two simulations with different seeds were started for a total 24 independent runs of 0.5 μs each (Supporting Information Table S1). Each 0.5 μs run took about 1 week on 64 cores of Xeon 5500-series processors.

Control simulations were carried out with the tetrapeptide Kac-Gly-Gly-Kac to investigate potential differences between the isolated Kac and histone tails. As in the runs with Kac, the tetrapeptide was capped with acetyl and *N*-methyl-amide groups at the N-terminus and C-terminus, respectively. The coordinates of the tetrapeptide and of the bromodomain BRD4(1) were taken from the PDB structure 3UVW. For CREBBP, the tetrapeptide was placed into its bindig site after structural overlap of the CREBBP coordinates (PDB entry code 3P1C) onto the ones of BRD4(1). From each of these two starting structures, two independent 1 μs runs were performed with the CHARMM PARAM27 force field. In addition, two independent 1 μs simulations with the AMBER force field were carried out for each of the two bromodomains. An overview of all simulations is given in Supporting Information Table S1.

Ligand Parametrization. For the simulations with the CHARMM PARAM27 force field, Kac was parametrized using the CHARMM general force field[41] for small molecules which is compatible with CHARMM PARAM27. For the simulations performed with the AMBER99 SB[27] force field, the parameters of the Kac side chains were taken from a previous computational study.[42]

Clustering and Cut-Based Free Energy Profile. The 16 MD trajectories with one or more binding events were used for further analysis (a total of 400 000 snapshots from 8 μs of MD). The snapshots were aligned structurally to the first frame of one of the MD trajectories using only the C_α atoms of the helical segments of TAF1(2). Subsequently, the snapshots were clustered using a tree-based algorithm[43] implemented in CAMPARI.[44] As a measure of pairwise similarity, the root-mean-square deviation (RMSD) of the acetyl-lysine side chain heavy atoms including C_α was computed without further alignment. The clustering was carried out with eight levels of the tree using 8 and 1 Å as RMSD thresholds for the coarsest and finest clustering level, respectively. This led to a total number of 149 024 clusters 126 552 of which contained only one snapshot (called singletons). The number of singletons seems large, but it has to be noted that in the unbound state, where the peptide moves freely around the bromodomain, the RMSD between two relative positions of Kac is almost always larger than 1 Å and thus the unbound state consists mainly of singletons. Other clustering methods which are able to partially solve the problem of the singletons were also tested but led to worse kinetic separation of the different states. Upon clustering, a free energy profile was computed using a method based on the equilibrium kinetic network which preserves the free energy barriers.[23,24,45] This approach emulates the cuts in flow-networks, and the computed profile is therefore named cut-based free energy profile. The nodes and links of the equilibrium kinetic network are the clusters (determined by the tree-based algorithm as explained above) and the direct transitions between them sampled along the MD runs, respectively.[46] For each node, the set of nodes is partitioned into two groups *A* and *B* using the mean first passage time (MFPT) to the reference node as an order parameter. The free energy is related to the maximum flow between sets *A* and *B* and calculated as $G = -kT\ \ln(Z_{AB}/Z)$, where Z_{AB}/Z is the relative partition function which represents the statistical weight of the transitions between sets *A* and *B* (see refs 23 and 24 for details). The result is a one-dimensional profile along the reaction coordinate Z_A/Z (i.e., the relative partition function representing the statistical weight of set *A*) which preserves the barrier height between the free energy basins. The evaluations of cut-based free energy profiles, interatomic distances, and dihedral angles were performed with the simulation analysis package WORDOM.[47,48]

Identification of Basins. The free energy basins were isolated from the cut-based free energy profile using a previously published procedure.[24] First the most populated cluster is used as reference to isolate the first basin which consists of the set of clusters on the left side of the first barrier on the profile. Due to possible overlap on the right side of the barrier (i.e., similar kinetic distance from two or more basins to the reference cluster), the isolation of each basin requires a separate profile[24] (Supporting Information Figures S7−S10). In other words, the procedure used for the first basin is iterated for the second basin which is isolated using as a reference cluster the largest cluster that is not in the first basin. Clusters on top of barriers are not assigned to any basin to avoid spurious fluctuations, i.e., shortcuts along the trajectories. Transitions between basins were calculated from the MD trajectories when the system passed from one basin to another either by a direct transition or via the not assigned snapshots (the barrier region).

Analysis of Hydrogen Bonds. Hydrogen bonds were analyzed using GROMACS.[34] A hydrogen bond between a donor and an acceptor atom was considered to be formed if the donor−acceptor distance was lower than 3.5 Å and the acceptor−donor−hydrogen angle lower than 30°.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Simulation protocol (Table S1); transition matrix computed after splitting of the N-binding mode according to the χ_2 dihedral angle of the conserved Asn (Table S2); time series of key interatomic distances for all the performed simulations (Figures S1−S4); time series of the root-mean-square displacement from the 3P1C crystal structure along a 500 ns simulation segment (Figure S3, bottom); time series of the root-mean-square displacement of Kac-Gly-Gly-Kac in complex with BRD4(1) and CREBBP from the 3UVW crystal structure (Figure S5); structural overlap of the representative snapshots of P- and N-binding mode with the structure of the complex between BRD4(1) and inhibitor reported in 4HXK (Figure S6); geometric annotation and cut-based free energy profile projections with representative nodes of the P-binding mode, N-binding mode, P/N-intermediate, and unbound state as reference nodes (Figure S7−S10); Movies S1 and S2 as mentioned in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: caflisch@bioc.uzh.ch. Phone: (+41 44) 635 55 21. Fax: (+41 44) 635 68 62.

Author Contributions

‡A.M. and S.S.: These authors contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Drs. Danzhi Huang and Andreas Vitalis for interesting discussions. We also thank Armin Widmer for continuous support with the program WITNOTP. The simulations were carried out on the Schrödinger cluster at the University of Zurich. This work was supported by grants of the Swiss National Science Foundation and the Swiss Cancer League (Krebsliga).

REFERENCES

(1) Zeng, L.; Zhou, M.-M. *FEBS Lett.* **2002**, *513*, 124−128.

(2) Filippakopoulos, P.; Knapp, S. *FEBS Lett.* **2012**, *586*, 2692−2704.

(3) Arrowsmith, C. H.; Bountra, C.; Fish, P. V.; Lee, K.; Schapira, M. *Nat. Rev. Drug. Discov.* **2012**, *11*, 384−400.

(4) Filippakopoulos, P.; et al. *Nature* **2010**, *468*, 1067−1073.

(5) Jenuwein, T.; Allis, C. D. *Science* **2001**, *293*, 1074−1080.

(6) Zuber, J.; et al. *Nature* **2011**, *478*, 524−528.

(7) Dawson, M. A.; Kouzarides, T.; Huntly, B. J. *N. Engl. J. Med.* **2012**, *367*, 647−657.

(8) Hewings, D. S.; Rooney, T. P. C.; Jennings, L. E.; Hay, D. A.; Schofield, C. J.; Brennan, P. E.; Knapp, S.; Conway, S. J. *J. Med. Chem.* **2012**, *55*, 9393−9413.

(9) Furdas, S. D.; Carlino, L.; Sippl, W.; Jung, M. *Med. Chem. Commun.* **2012**, *3*, 123−134.

(10) Dhalluin, C.; Carlson, J. E.; Zeng, L.; He, C.; Aggarwal, A. K.; Zhou, M.-M. *Nature* **1999**, *399*, 491−496.

(11) Mujtaba, S.; Zeng, L.; Zhou, M. *Oncogene* **2007**, *26*, 5521−5527.

(12) Filippakopoulos, P.; Picaud, S.; Mangos, M.; Keates, T.; Lambert, J.-P.; Barsyte-Lovejoy, D.; Felletar, I.; Volkmer, R.; Müller, S.; Pawson, T.; Gingras, A.-C.; Arrowsmith, C. H.; Knapp, S. *Cell* **2012**, *149*, 214−231.

(13) Vidler, L. R.; Brown, N.; Knapp, S.; Hoelder, S. *J. Med. Chem.* **2012**, *55*, 7346−7359.

(14) Owen, D. J.; Ornaghi, P.; Yang, J.-C.; Lowe, N.; Evans, P. R.; Ballario, P.; Neuhaus, D.; Filetici, P.; Travers, A. A. *EMBO J.* **2000**, *19*, 6141−6149.

(15) Huang, D.; Caflisch, A. *PLoS. Comput. Biol.* **2011**, *7*, e1002002.

(16) Seco, J.; Luque, F. J.; Barril, X. *J. Med. Chem.* **2009**, *52*, 2363−71.

(17) Guvench, O.; MacKerell, A. D. J. *PLoS. Comput. Biol.* **2009**, *5*, e1000435.

(18) Basse, N.; Kaar, J. L.; Settanni, G.; Joerger, A. C.; Rutherford, T. J.; Fersht, A. R. *Chem. Biol.* **2010**, *17*, 46−56.

(19) Pietrucci, F.; Marinelli, F.; Carloni, P.; Laio, A. *J. Am. Chem. Soc.* **2009**, *131*, 11811−11818.

(20) Buch, I.; Giorgino, T.; De Fabritiis, G. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 10184−10189.

(21) Huang, D.; Caflisch, A. *ChemMedChem* **2011**, *6*, 1578−1580.

(22) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 13118−13123.

(23) Krivov, S. V.; Karplus, M. *J. Phys. Chem. B* **2006**, *110*, 12689−12698.

(24) Krivov, S. V.; Muff, S.; Caflisch, A.; Karplus, M. *J. Phys. Chem. B* **2008**, *112*, 8701−8714.

(25) Steiner, S.; Magno, A.; Huang, D.; Caflisch, A. *FEBS Lett.* **2013**, *587*, 2158−2163.

(26) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III *J. Comput. Chem.* **2004**, *25*, 1400−1415.

(27) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.

(28) Chung, C. W.; et al. *J. Med. Chem.* **2011**, *54*, 3827−3838.

(29) Chung, C. W.; Dean, A. W.; Woolven, J. M.; Bamborough, P. *J. Med. Chem.* **2012**, *55*, 576−586.

(30) Moriniere, J.; Rousseaux, S.; Steuerwald, U.; Soler-Lopez, M.; Curtet, S.; Vitte, A.-L.; Govin, J.; Gaucher, J.; Sadoul, K.; Hart, D. J.; Krijgsveld, J.; Khochbin, S.; Muller, C. W.; Petosa, C. *Nature* **2009**, *461*, 664−668.

(31) Zhao, L.; et al. *J. Med. Chem.* **2013**, *56*, 3833−3851.

(32) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(33) Brooks, B. R.; et al. *J. Comput. Chem.* **2009**, *30*, 1545−1614.

(34) Spoel, D. V. D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701−1718.

(35) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L. J.; J., D. E.; et al. *J. Phys. Chem. B* **1998**, *102*, 35863616.

(36) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926935.

(37) Darden, T.; York, D.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *98*, 10089.

(38) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.

(39) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(40) *The PyMOL Molecular Graphics System*, version 1.3r1; Schrödinger, LLC, New York, 2010.

(41) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. *J. Comput. Chem.* **2010**, *31*, 671−690.

(42) Liu, H.; Duan, Y. *Biophys. J.* **2008**, *94*, 4579−4585.

(43) Vitalis, A.; Caflisch, A. *J. Chem. Theory Comput.* **2012**, *8*, 1108−1120.

(44) Vitalis, A.; Steffen, A.and Lyle, N.; Mao, A. H.; Pappu, R. V. *CAMPARI v1.0.* http://sourceforge.net/projects/campari (accessed April 22, 2013).

(45) Krivov, S. V.; Karplus, M. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 13841−13846.

(46) Rao, F.; Caflisch, A. *J. Mol. Biol.* **2004**, *342*, 299−306.

(47) Seeber, M.; Cecchini, M.; Rao, F.; Settanni, G.; Caflisch, A. *Bioinformatics* **2007**, *23*, 2625−2627.

(48) Seeber, M.; Felline, A.; Raimondi, F.; Muff, S.; Friedman, R.; Rao, F.; Caflisch, A.; Fanelli, F. *J. Comput. Chem.* **2011**, *32*, 1183−1194.

Mechanism and kinetics of acetyl-lysine binding to bromodomains

SUPPLEMENTARY INFORMATION

A. Magno[a‡], S. Steiner[a‡], and A. Caflisch[a*]

[a]Department of Biochemistry
University of Zürich, Winterthurerstrasse 190
CH-8057 Zürich, Switzerland
Phone: (+41 44) 635 55 21, FAX: (+41 44) 635 68 62
email: caflisch@bioc.uzh.ch

[*]Corresponding author

[‡] These authors contributed equally to this work

keywords: molecular dynamics, post-translational modifications, histones, epigenetic marks, acetylated lysine, cut-based free energy profile

July 4, 2013

Bromodomain (PDB code)	Ligand	Force Field	Starting structure	No. of runs	Total sampling [μs]
TAF1(2) (3UV4)	Kac	CHARMM	Unbound, Asn1604 χ_2 up	12	6
TAF1(2) (3UV4)	Kac	CHARMM	Unbound, Asn1604 χ_2 down	12	6
TAF1(2)	Kac	AMBER	P-binding from CHARMM	2	2
TAF1(2)	Kac	AMBER	N-binding from CHARMM	2	2
BRD4(1) (3UVW)	Tetrapeptide	CHARMM	X-Ray	2	2
CREBBP (3P1C)	Tetrapeptide	CHARMM	Overlap on 3UVW	2	2
BRD4(1)	Tetrapeptide	AMBER	from CHARMM run	2	2
CREBBP	Tetrapeptide	AMBER	from CHARMM run	2	2

Table S1: Simulations performed.

Transitions from \ to	**Unbound**	**P-binding**	**N-binding DOWN**	**N-binding UP**	**P/N-Intermediate**
Unbound	-	1	2	12	5
P-binding	2	-	119	10	310
N-binding DOWN	2	118	-	18	4
N-binding UP	2	17	18	-	13
P/N-Intermediate	4	312	6	10	-

Table S2: The number of transitions between the different basins after splitting of the N-binding basin into two sub-basins characterized by different values of Asn1604 χ_2.

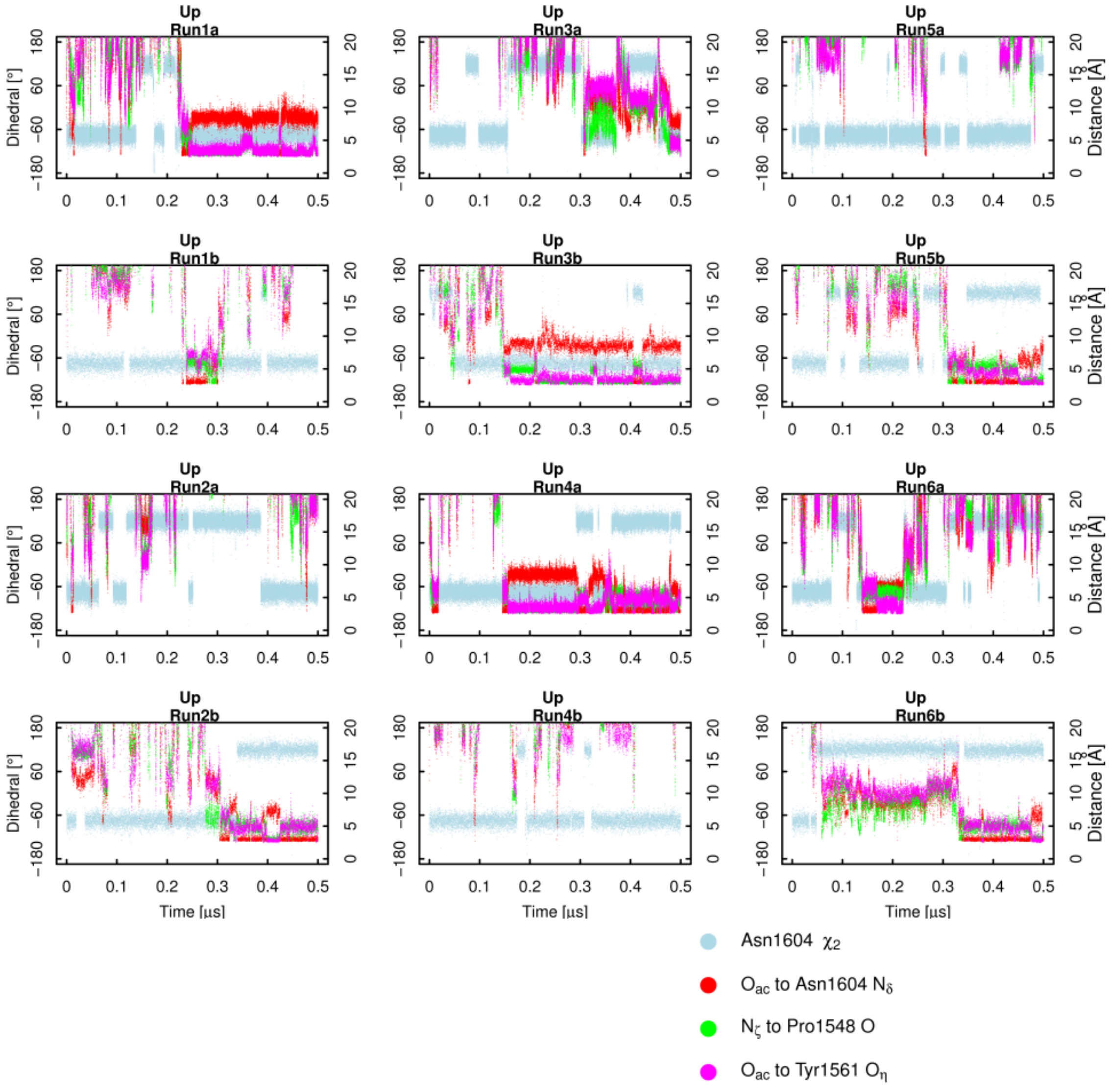


Figure S1: Time series of the interatomic distances between donor and acceptor atoms of three intermolecular key interactions (red, green, and magenta, y-axis on the right) together with the time series of the χ_2 dihedral angle of the conserved Asn1604 (cyan, y-axis on the left) for the 12 individual runs started with the N_δ atom of Asn1604 pointing towards the solvent (Up). The letters *a* and *b* distinguish two runs started from the same starting structure using different seeds.

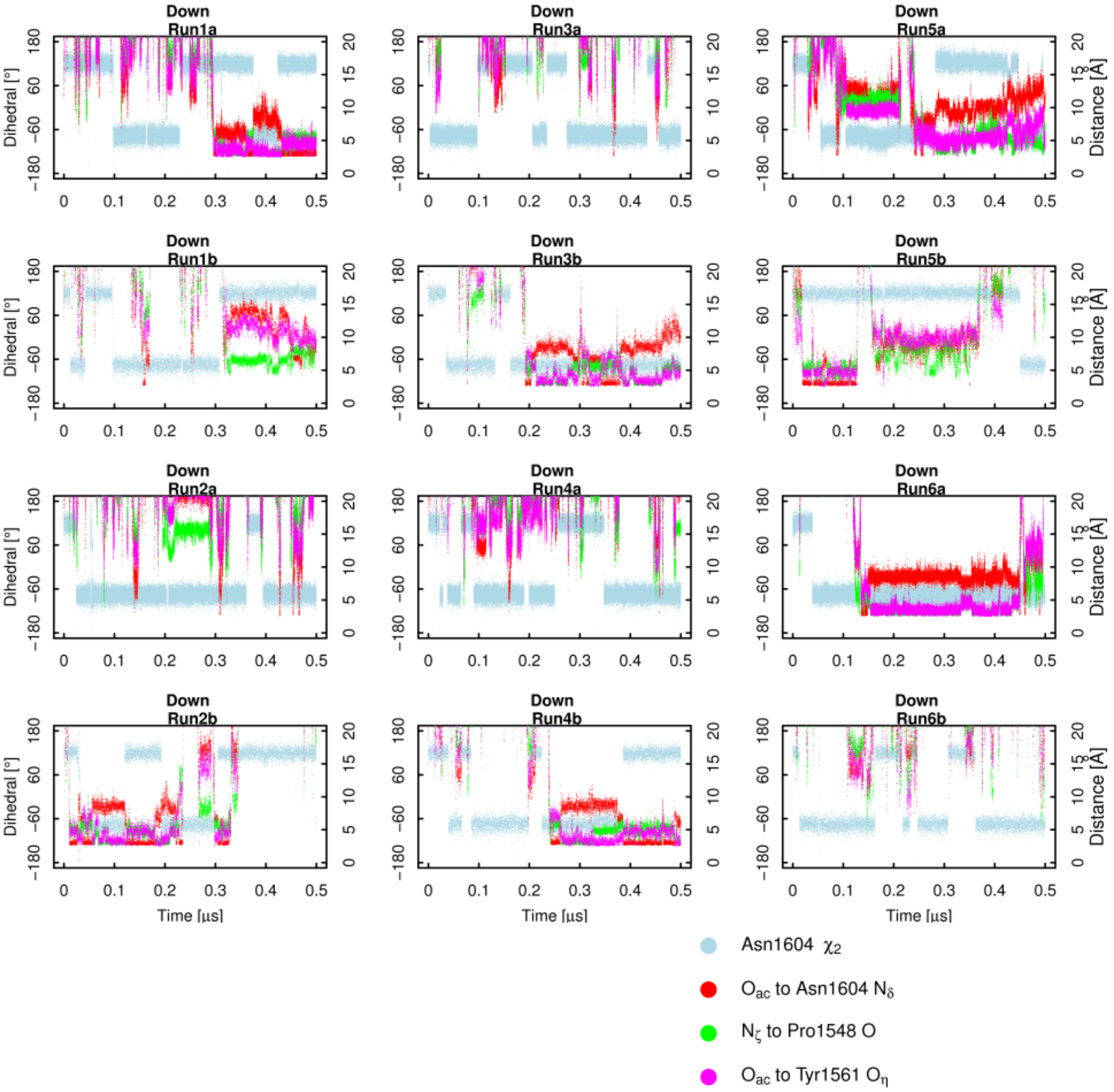


Figure S2: Same as in Figure S1 for the 12 individual runs started with the N_δ atom of Asn1604 pointing towards the binding site bottom (Down).

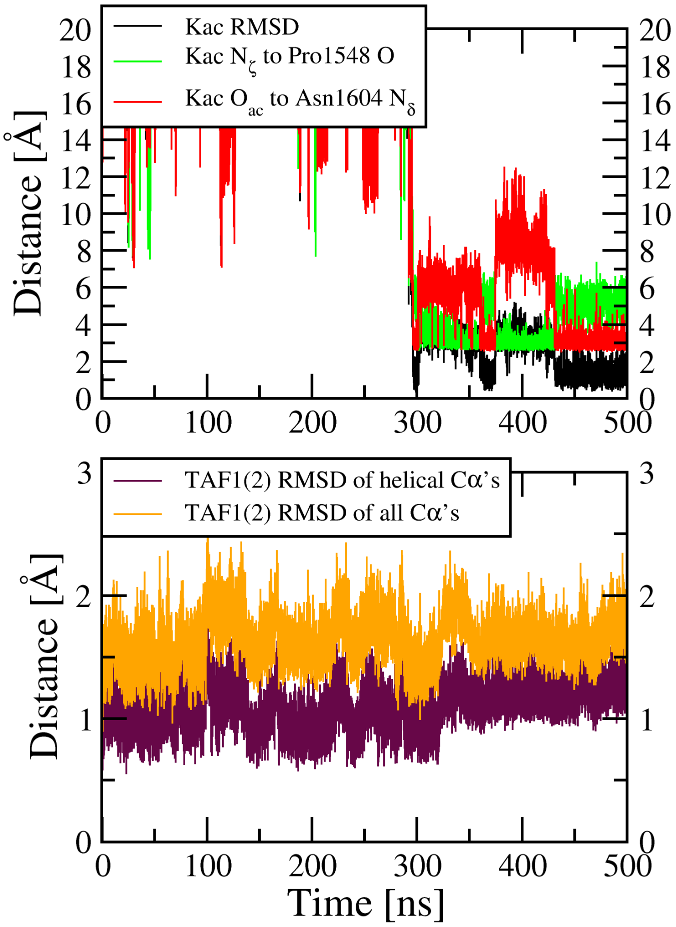


Figure S3: (Top) Time series of the RMSD from the X-ray structure (black) together with the distances that characterize the hydrogen bond observed in the crystal structures (red) and the hydrogen bond of the more buried binding mode revealed by the present MD study (green). The RMSD of the Kac side chain atoms was calculated upon optimal overlap of the C_α atoms in the four helices. The reference structure for the RMSD calculation was the crystal structure of CREBBP in complex with Kac (PDB code 3P1C) as there is no crystal structure of TAF1(2) with Kac or a histone tail peptide. (Bottom) Time series of the RMSD from the equilibrated conformation of TAF1(2) C_α atoms (orange) and C_α atoms in helices only (maroon). These time series show that upon Kac binding the bromodomain structure becomes slightly less flexible but does not undergo any relevant conformational change.

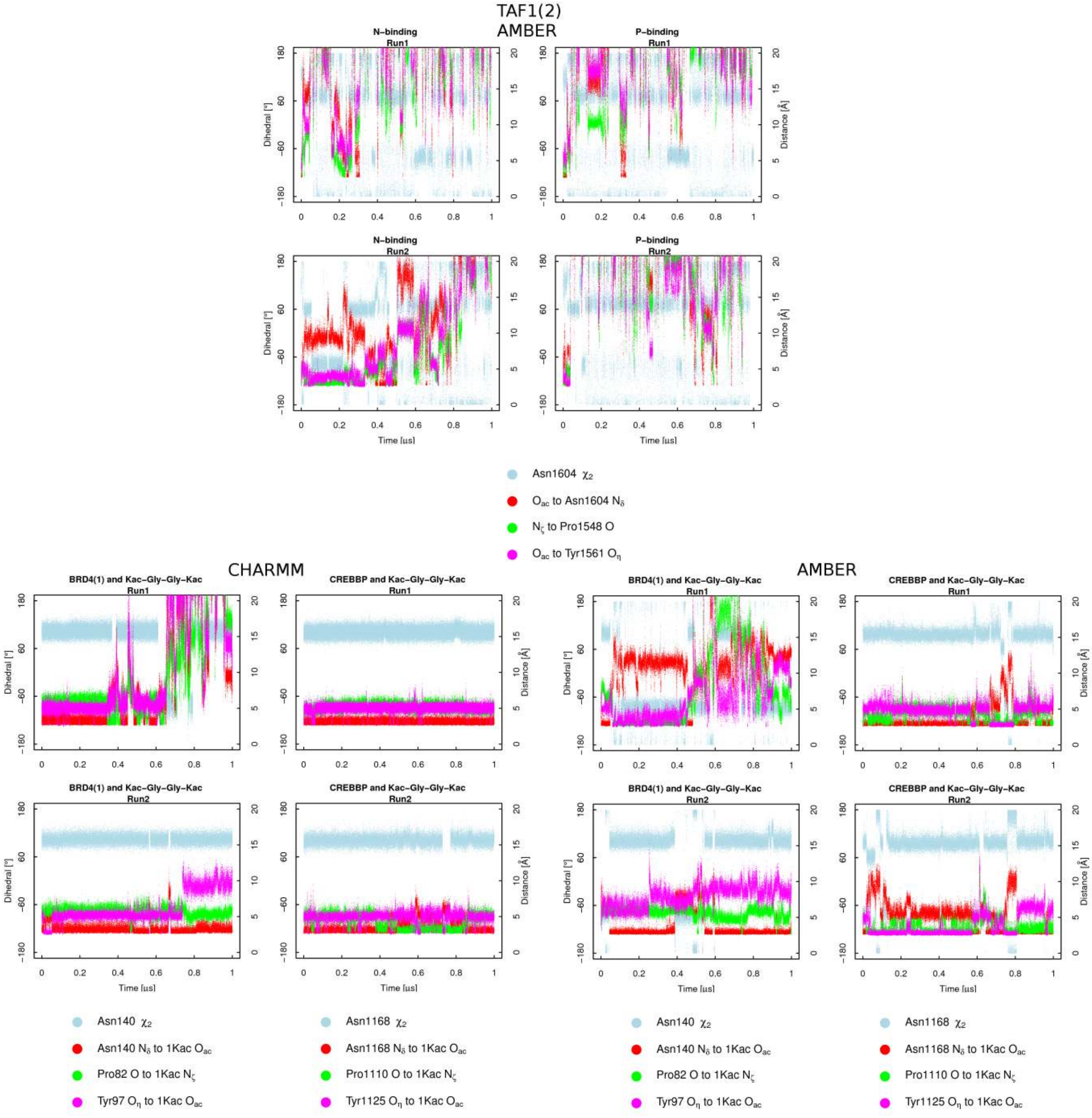


Figure S4: Control simulations with the AMBER force field and/or the tetrapeptide Kac-Gly-Gly-Kac. (Top) Time series of the interatomic distances between donor and acceptor atoms of three intermolecular key interactions (red, green, and magenta, y-axis on the right) together with the time series of the χ_2 dihedral angle of the conserved Asn in the binding site (cyan, y-axis on the left) for the control simulations of TAF1(2) and Kac with AMBER force field. (Bottom) Time series of interatomic distances between the N-terminal Kac of the tetrapeptide Kac-Gly-Gly-Kac and BRD4(1) or CREBBP for the simulations with the CHARMM (bottom left) or AMBER (bottom right) force field.

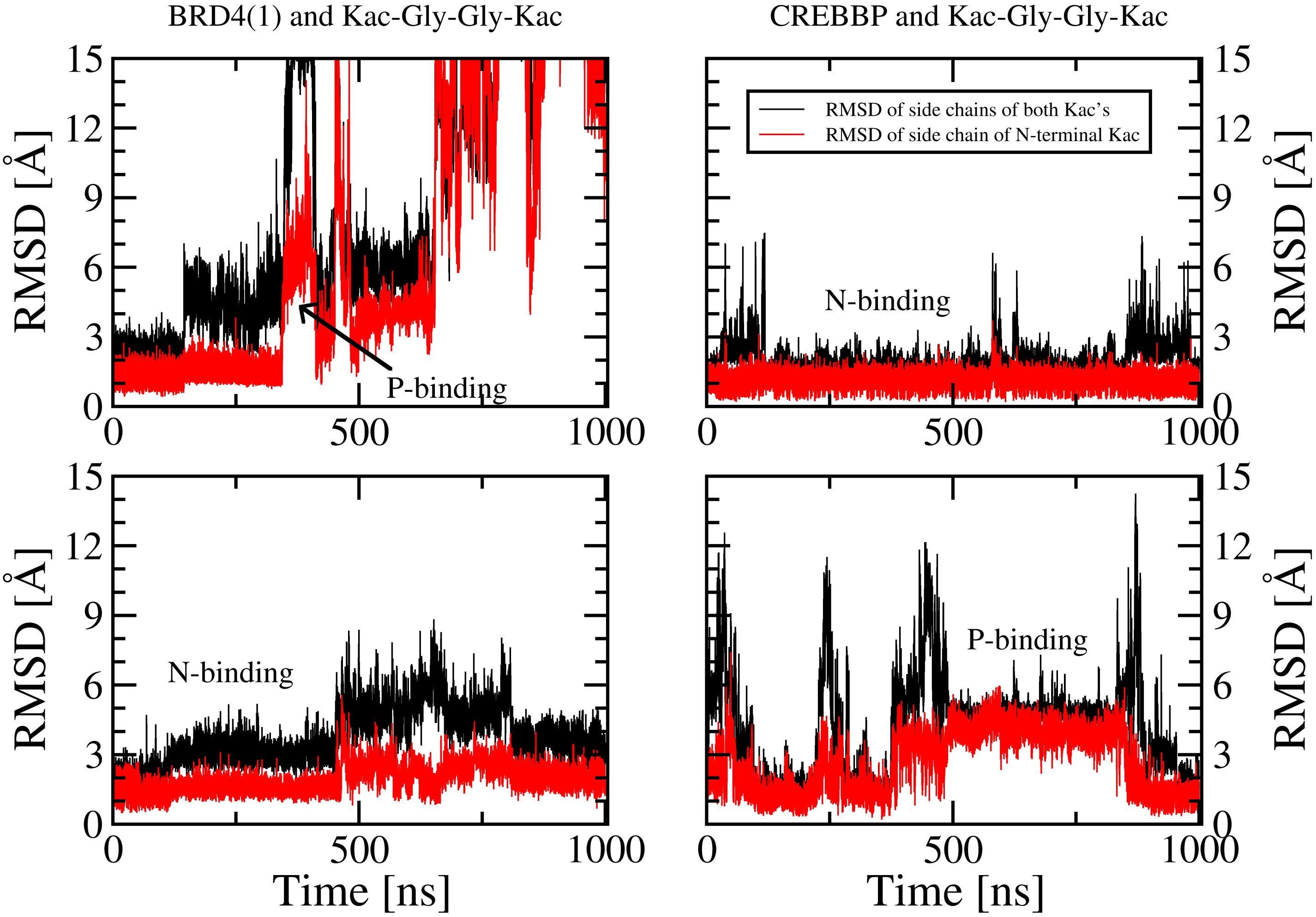


Figure S5: Control simulations with the CHARMM force field and the tetrapeptide Kac-Gly-Gly-Kac. Time series of the RMSD of the side chain atoms of the two Kac's (black) and only the N-terminal Kac (red) for two independent runs each with BRD4(1) (left) and CREBBP (right). The RMSD was calculated upon optimal overlap of the C_α atoms in the four helices of the bromodomain. The reference structure for the RMSD calculation was the crystal structure of BRD4(1) in the complex with Kac-Gly-Gly-Kac from the histone 4 peptide H4K5acK8ac (PDB code 3UVW).

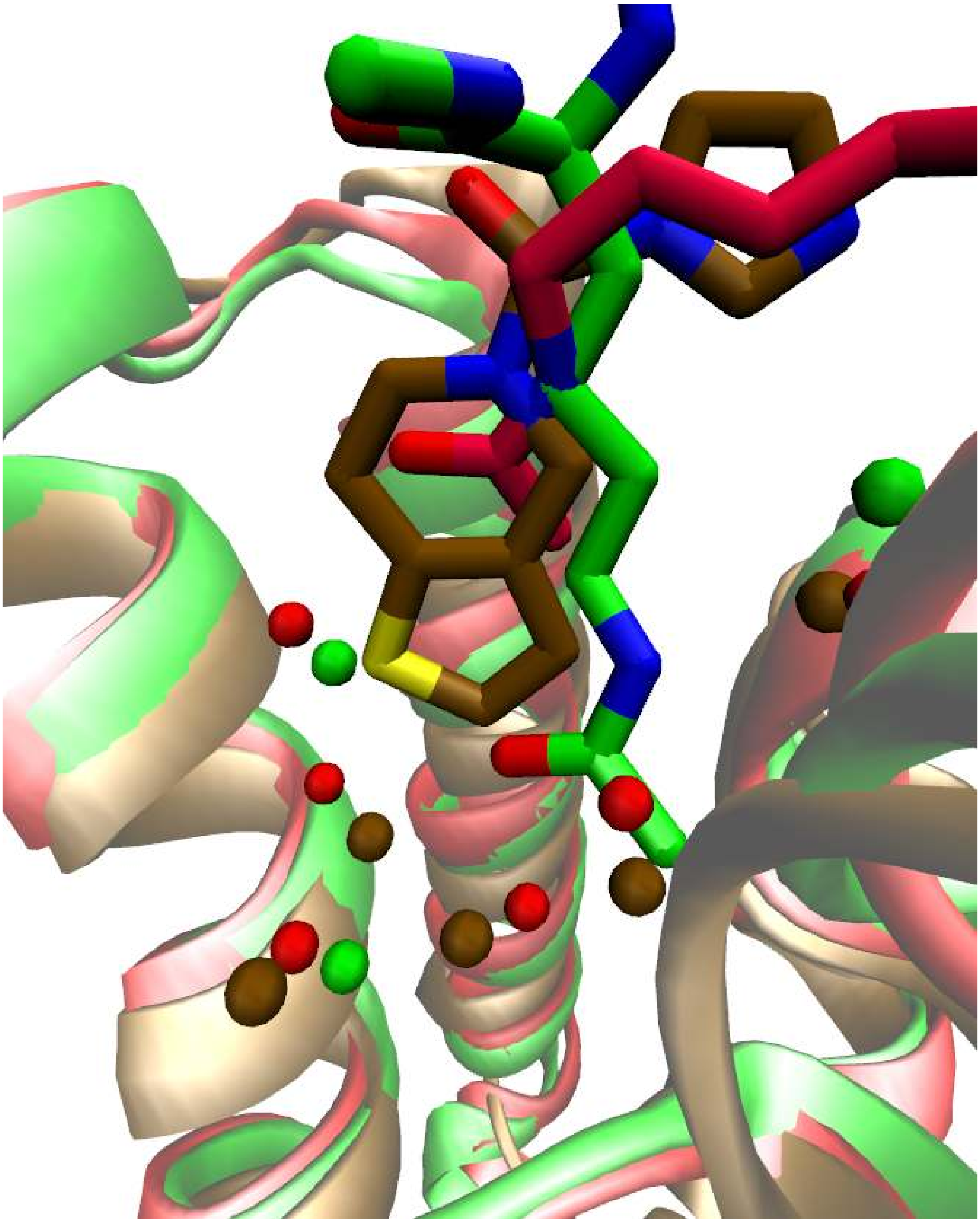


Figure S6: Small molecule inhibitors of brodomains can penetrate deeply into the acetyl-lysine binding pocket similarly to the P-binding mode of Kac observed in the MD simulations. Representative snapshots of acetyl-lysine in the P-binding mode (green) and N-binding mode (red) are structurally aligned, using only the C_α atoms of the bromodomain helices, to the crystal structure of BRD4(1) in complex with a recently published inhibitor (brown, PDB code 4HXK). The bromodomain backbone, Kac/inhibitor, and structured water molecules are shown as ribbons, sticks and spheres, respectively.

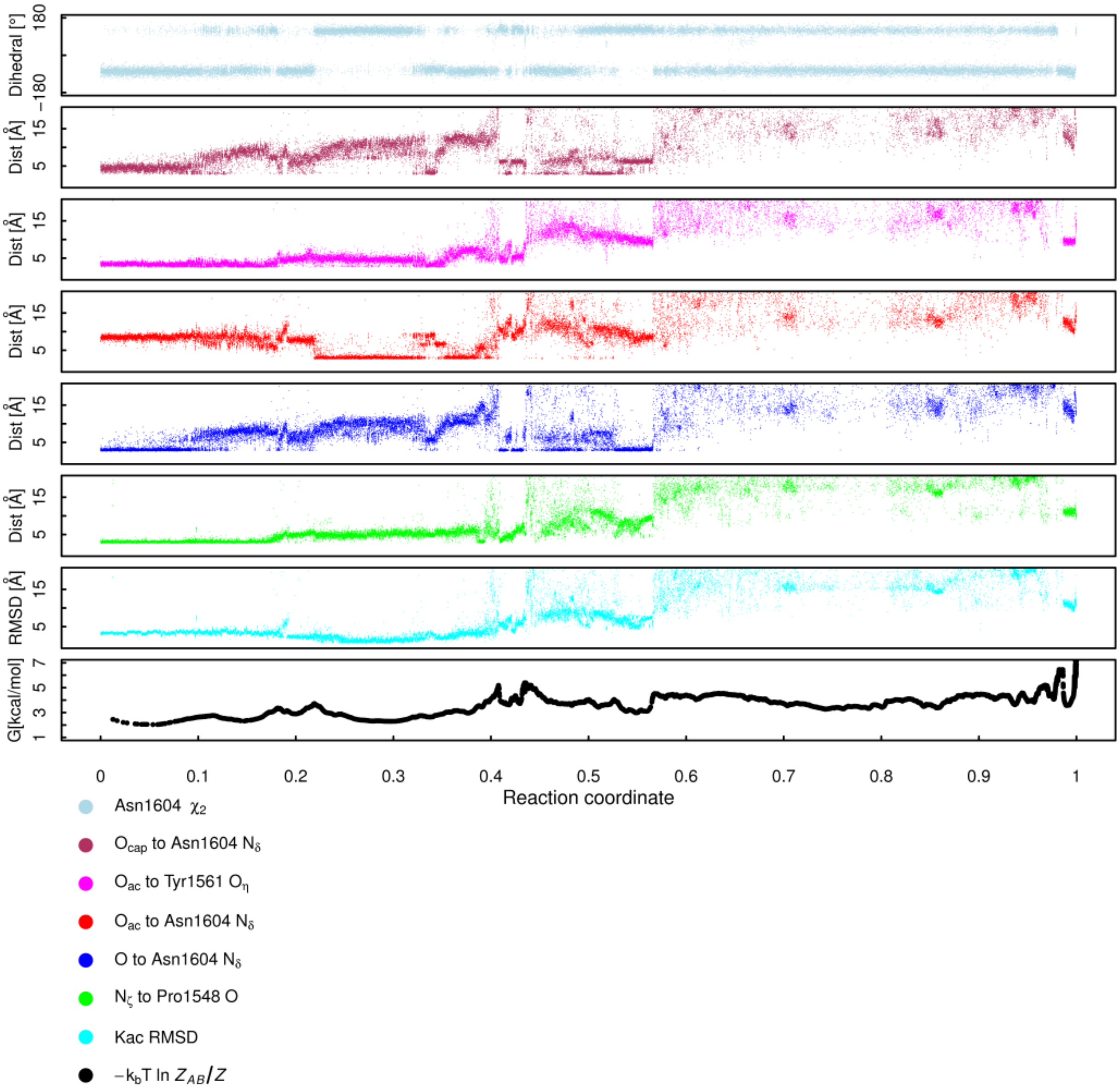


Figure S7: Geometric annotation and cFEP (bottom panel) with the P-binding mode representative as reference node. The RMSD (second panel from the bottom) of the Kac atoms with respect to the X-ray structure reported in 3P1C was calculated after overlap of the TAF1(2) C_α atoms to the corresponding atoms of CREBBP in 3P1C.

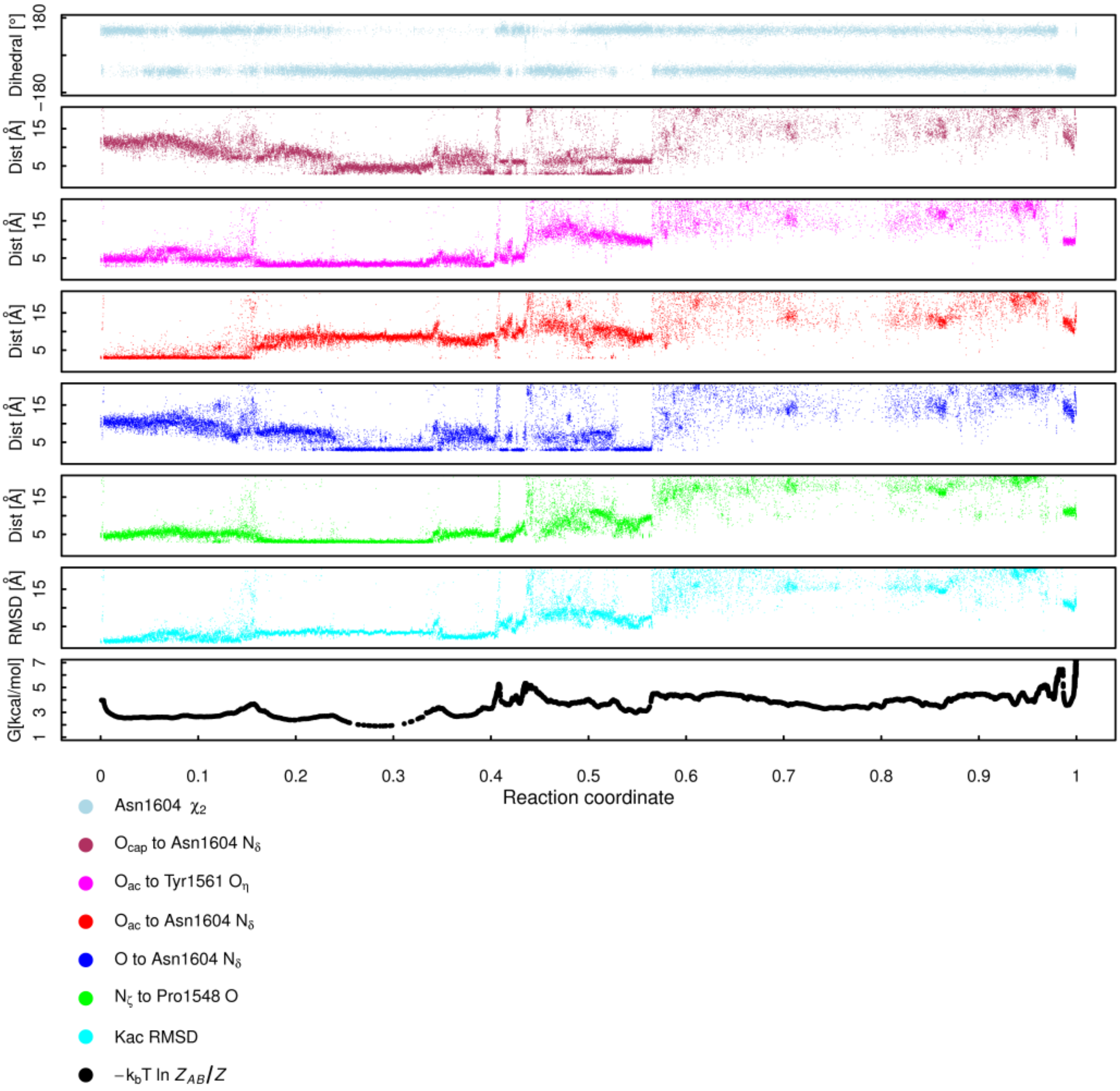


Figure S8: Geometric annotation and cFEP (bottom panel) with the N-binding mode representative as reference node. For further details see legend of Figure S7.

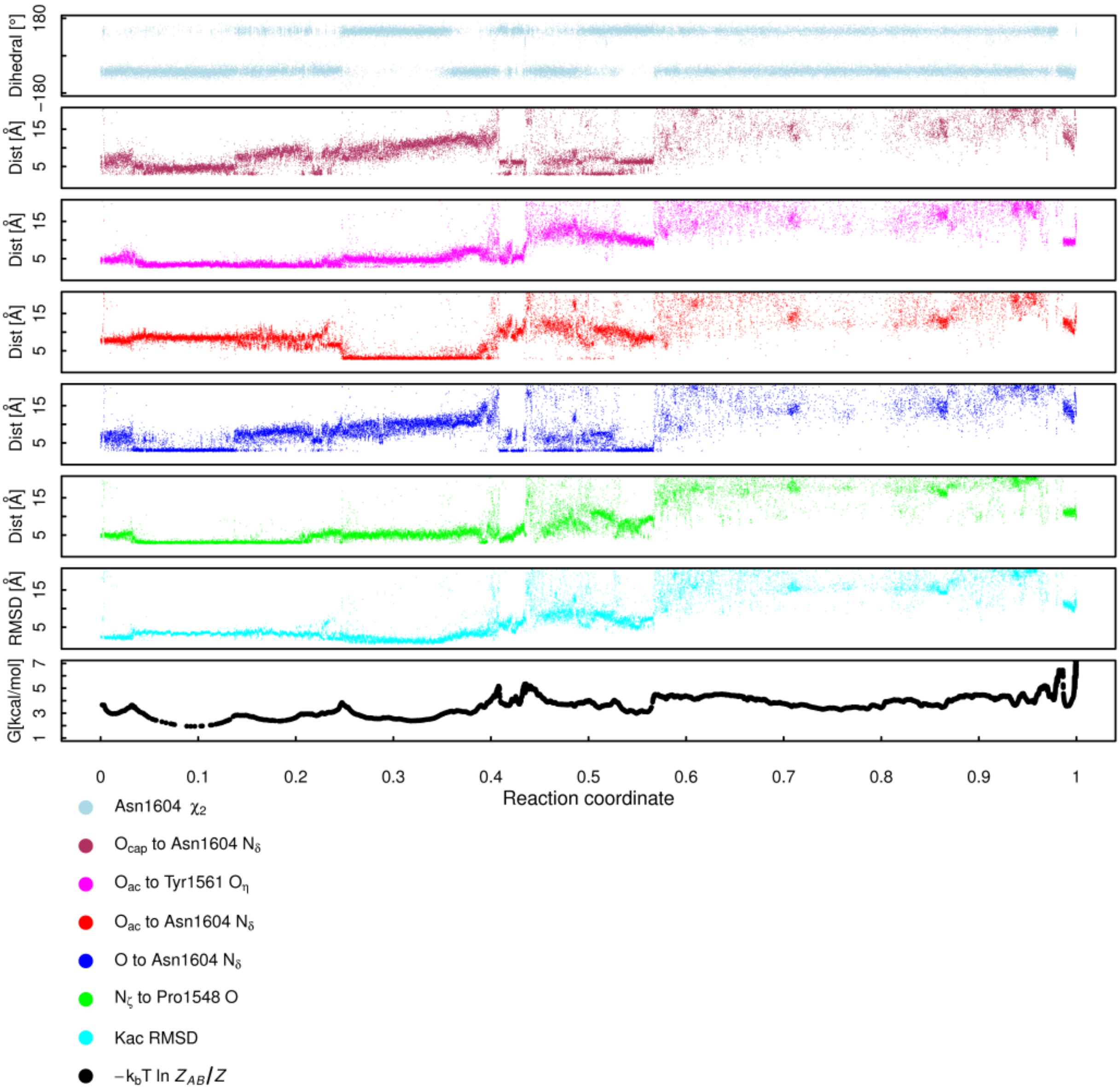


Figure S9: Geometric annotation and cFEP (bottom panel) with the P/N-Intermediate representative as reference node. For further details see legend of Figure S7.

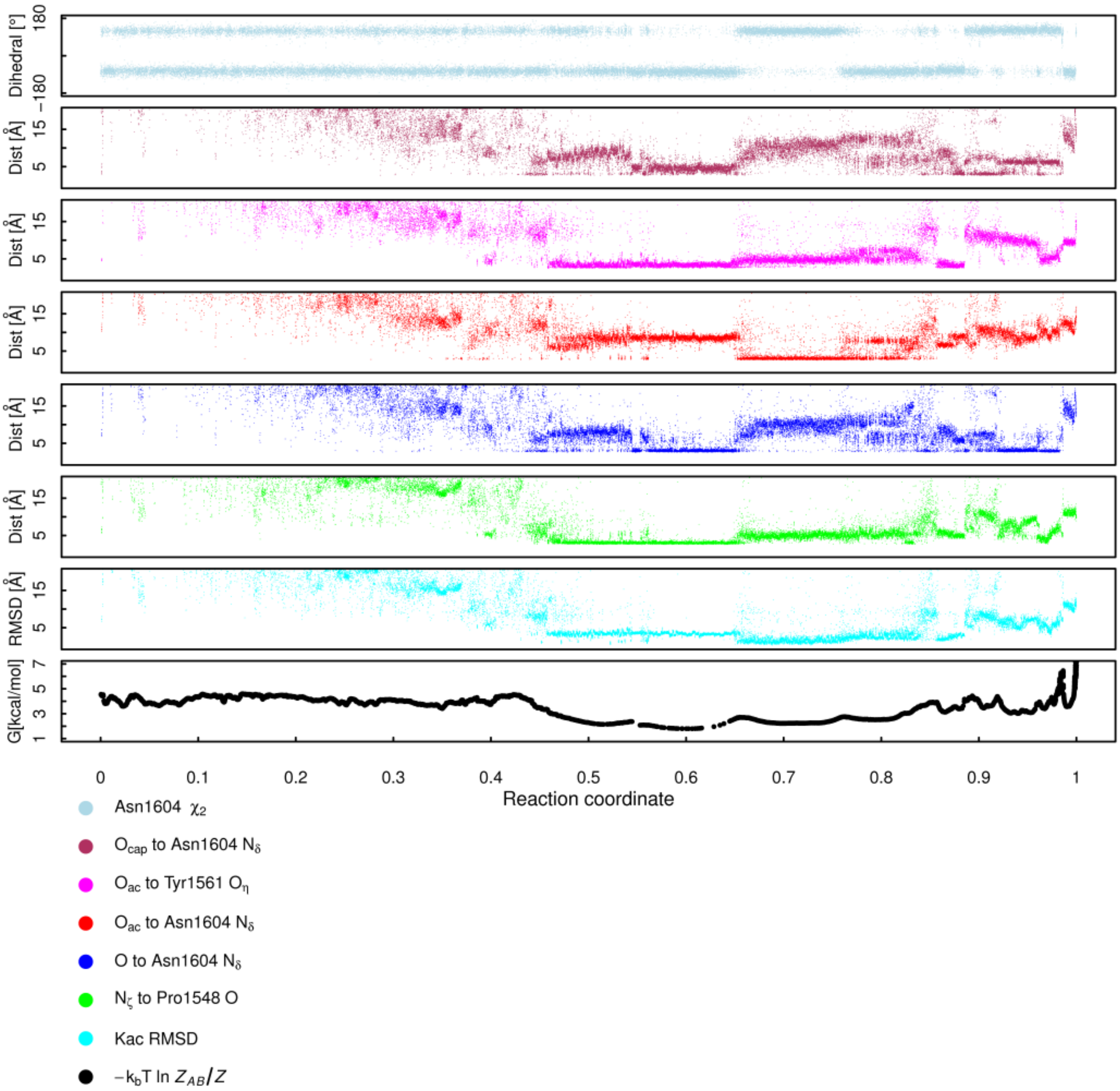


Figure S10: Geometric annotation and cFEP (bottom panel) from the unbound state. For further details see legend of Figure S7.

Chapter 8

Conclusions

In order to investigate protein-protein assembly we have studied several aspects of this process by MD simulations. First of all, it is important to note that already at the starting point of the problem, i.e., the protein binding site, we are presented with a number of interesting phenomena which are necessary to be studied in detail in order to get a further understanding of ligand binding by proteins. Our research provides an example of a protein (PDZ3 of PSD-95) which can adopt three conformations in the native ligand-free state while in the bound state only one of these conformations have been observed. Among other structural properties, the three conformations differ by how "open" the binding site is. The simulations indicate that ligand binding by PDZ3 follows a conformational selection mechanism. Most interestingly, also members of the bromodomain family show varied plasticity of the binding site in the native ligand-free state. They differ not only by different orientations of key residues responsible for binding its natural ligand acetyl-lysine, but also in solvent accessibility and even an completely obstructed conformation of the binding site has been observed, which is hypothesized to be unavailable for ligand-binding.

Additionally, investigation of the unbinding process of a 6-residue peptide ligand from a PDZ domain (PDZ2 of hPTP1E) which is bound by extending a β-hairpin of the domain to a three-stranded β-sheet, suggests that the intermolecular backbone interactions with C-terminal residues of the ligand are particularly stable compared to backbone interactions with N-terminal ligand residues. Consistently, a metastable state different from the native bound state has been observed where the ligand is bound to the protein only via its C-terminal residues.

The conformational change a protein undergoes upon binding of a peptide ligand has been studied by covalently linking an azobenzene-photoswitch across the binding groove of PDZ2 of hPTP1E and imitating via its isomerization the two states of the protein. The transition was investigated by ultrafast IR spectroscopy, NMR spectroscopy and MD simulations. The latter showed that water friction plays a critical role and that the propagation of the perturbation of the water might be an example for allosteric signaling.

Finally, the simulations of a bromodomain (TAF1(2)) together with a peptide mimicking its natural ligand acetyl-lysine, show that thanks to fast technical development by now MD simulations are able to capture multiple binding events of small peptidic ligands. Furthermore, the simulations allow us not only to extract the mechanistic and kinetic details of binding, but in this case also reveal an alternative binding mode, which has not been experimentally observed up to now. Additionally, in this alternative binding mode several of the structured water molecules inside the binding pocket are replaced by the ligand. This is an observation that is in line with the studies of bromodomains and different cosolvent molecules, where we show that some of the structured water molecules can be displaced transiently by cosolvent.

In conclusion, MD simulations provide an excellent tool to study dynamical and structural aspects of binding unaccessible by the usual experimental techniques and thereby contribute important insights in the highly complex process of ligand binding by proteins.

Acknowledgements

I would like to thank Prof. Dr. Amedeo Caflisch for the possibility to work in his group and for his continuous support during these 4 years of PhD studies. Furthermore, I would like to express my gratitude towards Prof. Dr. Peter Hamm and the members of his group for involving me in their highly interesting research project.

Naturally, I thank also all my actual and former colleagues for fruitful and interesting discussions and help with computer issues every once in a while. The latter applies as well to the IT service of the Institute of Biochemistry.

Moreover, I am very grateful to my family and friends who have supported me during the good and bad times of the doctorate.

While I was a PhD student, a colleague and dear friend fell ill and passed away. He was an exceptional person, and I am grateful for having known him.

Concludendo, voglio ringraziare il mio miglior amico e compagno di vita per aver cura di me, sempre.

List of publications

Peptide binding to the PDZ3 domain by conformational selection.
Steiner, S. and Caflisch, A.
[*Proteins: Structure, Function and Bioinformatics*, 80(11): 2562-2572, 2012]

Kinetic response of a photo-perturbed allosteric protein.
Buchli, B., Waldauer, S.A., Walser, R., Donten, M., Pfister, R., Blöchliger, N., Steiner, S., Caflisch, A., Zerbe, O. and Hamm, P.
[*Proceedings of the National Academy of Sciences*, 110(29): 11725-11730, 2013]

Does bromodomain flexibility influence histone recognition?
Steiner, S., Magno, A., Huang, D. and Caflisch, A.
[*FEBS letters*, 587(14): 2158-2163, 2013]

Structured water molecules in bromodomains binding site can be displaced by cosolvent.
Huang, D., Rossini, E., Steiner, S. and Caflisch, A.
[Article in press at *ChemMedChem*]

Mechanism and kinetics of acetyl-lysine binding to bromodomains.
Magno, A.[‡], Steiner, S.[‡] and Caflisch, A.
[*Journal of Chemical Theory and Computation*, 9(9): 4225-4232, 2013]
[‡] These authors contributed equally to this work

Curriculum Vitae

Sandra Steiner

born 30th March 1985 in Zurich, Switzerland

from Schänis SG

Education

November 2009 - July 2013	
	PhD student in the group of Prof. Dr. Amedeo Caflisch, Department of Biochemistry, University of Zurich
September 2009	
	Master degree in Mathematics at the University of Zurich, Switzerland, grade 5.9
October 2004 - September 2009	
	Studies of Mathematics with specialization in Coding Theory and Cryptography at the University of Zurich, Switzerland
September 1997 - July 2003	
	High School Diploma at Literargynamsium Rämibühl in Zurich, Switzerland